Space Programs Summary No. 37-26, Volume IV

for the period February 1, 1964 to March 31, 1964

Supporting Research and Advanced Development

jpl

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

April 30, 1964

Space Programs Summary No. 37-26, Volume IV

for the period February 1, 1964 to March 31, 1964

Supporting Research and Advanced Development

JET PROPULSION LABORATORY
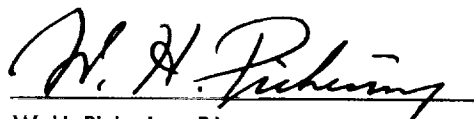CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

April 30, 1964

Preface

The *Space Programs Summary* is a six volume, bimonthly publication designed to report on JPL space exploration programs, and related supporting research and advanced development projects. The subtitles of all volumes of the *Space Programs Summary* are:

Vol. I.   The Lunar Program (Confidential)

Vol. II.   The Planetary–Interplanetary Program (Confidential)

Vol. III. The Deep Space Instrumentation Facility (Unclassified)

Vol. IV.   Supporting Research and Advanced Development (Unclassified)

Vol. V.   Supporting Research and Advanced Development (Confidential)

Vol. VI.   Space Exploration Programs and Space Sciences (Unclassified)

The *Space Programs Summary*, Volume VI consists of: an unclassified digest of appropriate material from Volumes I, II, and III; original presentation of the JPL Space Flight Operations Facility development progress; and a reprint of the space science instrumentation studies of Volumes I and II.

W. H. Pickering, *Director*
Jet Propulsion Laboratory

Contents

SYSTEMS DIVISION

GUIDANCE AND CONTROL DIVISION

Contents (cont'd)

Contents (cont'd)

Contents (cont'd)

Contents (cont'd)

SYSTEMS DIVISION

I. Systems Analysis

A. The Equations of Motion for the Plane Restricted Three-Body Problem

R. Broucke

1. Introduction

The first approximation which is generally used for the study of trajectories in the Earth-Moon system is the two-body problem. The Earth is used as the central body and the Moon is simply neglected in one part of the trajectory, while in another part of the trajectory the Moon becomes the central body and the Earth influence is neglected. But there is a much better mathematical model for the study of the flight mechanics in the Earth-Moon system, consisting of the restricted three-body problem with the Earth and Moon as main masses, and the satellite as a third body. We have now undertaken at JPL a systematic study of the restricted three-body problem, starting with the planar circular case. And, since the first step in such a study is the establishment of the equations of motion, we summarize here a dozen forms for these equations, using different coordinate systems. Since the circular restricted three-body problem is a conservative problem, we generally use the Lagrangian and Hamiltonian formalism in order to perform the transformations. The equations of motion are then derived directly from the Lagrangian or the Hamiltonian. For this reason we do not always give them explicitly. Our most important goal in deriving different forms of equations of motion is to arrive at equations which do not present the singularities $r_1 = 0$ or $r_2 = 0$, that is, which do not have vanishing denominators when one of the distances, Earth-satellite r_1, or Moon-satellite r_2, tends to zero. The last forms of equations we give in this summary all satisfy this condition. This question of removing the singularities is the so-called regularization.

2. The Restricted Three-Body Problem with Inertial and Synodic Coordinates

We give here the definition of the plane circular restricted three-body problem and the corresponding fundamental equations, using rectangular coordinate systems with a fixed orientation, or rotating about the center of mass.

The circular restricted three-body problem studies the motion of a satellite in the gravitation field of two masses which are rotating in their Keplerian motion around the common barycenter. We shall suppose here that the satellite's path is always in the plane of the motion of the two main masses. We are essentially interested in the mass ratio $m_1/m_2 = 81.30$, which is the approximate value of the Earth-Moon mass ratio. On the other hand, we take the total mass $m_1 + m_2$ as the mass unit and then we can set $m_1 = 1 - \mu$ and $m_2 = \mu$, with $\mu = 1/82.30$. We also take the constant length $|m_1 m_2|$ as the unit of length. In this way, the circles described by m_1 and m_2 around the center of mass have the respective radii $|Om_1| = \mu$ and $|Om_2| = 1 - \mu$. The unit of time has been chosen in such a way that the mean angular velocity n for m_1 and m_2 about the center of mass is $+1$. With this choice of units Kepler's third law shows us that the gravitation constant is $+1$:

$$n^2 |m_1 m_2|^3 = G(m_1 + m_2).$$

We can refer the problem to a fixed rectangular coordinate system with the origin at the center of mass. Let (ξ, η) be the coordinates of the satellite m_3 in this coordinate system. We can always take the coordinate system in such a way that at the instant $t = 0$, m_1 and m_2 are on the $O\xi$-axis, at the abscissa $-\mu$ and $1 - \mu$. At any time, the coordinates of the satellite are then

$$\xi_1 = -\mu \cos t, \qquad \xi_2 = (1 - \mu) \cos t,$$

$$\eta_1 = -\mu \sin t, \qquad \eta_2 = (1 - \mu) \sin t.$$

The distances r_1 and r_2 between the satellite m_3 and the two main masses m_1 and m_2 are given by

$$r_1^2 = (\xi - \xi_1)^2 + (\eta - \eta_1)^2,$$
$$r_2^2 = (\xi - \xi_2)^2 + (\eta - \eta_2)^2.$$

The differential equations for the satellite's motion are then

$$\frac{d^2\xi}{dt^2} = -(1 - \mu)\frac{\xi - \xi_1}{r_1^3} - \mu\frac{\xi - \xi_2}{r_2^3},$$

$$\frac{d^2\eta}{dt^2} = -(1 - \mu)\frac{\eta - \eta_1}{r_1^3} - \mu\frac{\eta - \eta_2}{r_2^3}.$$

We now introduce a coordinate system which is rotating about the center of mass with the uniform angular velocity $+1$. Let (x, y) be the coordinates of the satellite m_3 in this new rotating frame. We can always convert

from the fixed coordinate system to the rotating coordinate system by

$$\xi = x \cos t - y \sin t,$$

$$\eta = x \sin t + y \cos t.$$

The coordinates for m_1 and m_2 in this new system are

$$x_1 = -\mu, \qquad\qquad y_1 = 0,$$

$$x_2 = 1 - \mu, \qquad\qquad y_2 = 0.$$

The distances r_1 and r_2 are given by the expressions

$$r_1^2 = (x - x_1)^2 + y^2,$$

$$r_2^2 = (x - x_2)^2 + y^2,$$

and the new differential equations for the motion are

$$\frac{d^2x}{dt^2} - 2\frac{dy}{dt} - x = -(1 - \mu)\frac{x - x_1}{r_1^3} - \mu\frac{x - x_2}{r_2^3},$$

$$\frac{d^2y}{dt^2} + 2\frac{dx}{dt} - y = -(1 - \mu)\frac{y}{r_1^3} - \mu\frac{y}{r_2^3}.$$

The equations of motion in the fixed coordinate system (ξ, η) can also be derived from the Lagrangian

$$L = \frac{1}{2}(\xi'^2 + \eta'^2) + \left(\frac{1 - \mu}{r_1} + \frac{\mu}{r_2}\right).$$

By defining the canonical moments (p_ξ, p_η) associated with coordinates (ξ, η) by $p_\xi = \xi'$, $p_\eta = \eta'$, we can derive the Hamiltonian

$$H = \frac{1}{2}(p_\xi^2 + p_\eta^2) - \left(\frac{1 - \mu}{r_1} + \frac{\mu}{r_2}\right) = \frac{1}{2}(p_\xi^2 + p_\eta^2) + V.$$

In the rotating coordinate system (x, y), the Lagrangian takes the form

$$L = \frac{1}{2}(x'^2 + y'^2) + \frac{1}{2}(x^2 + y^2)$$

$$+ (xy' - yx') + \frac{1 - \mu}{r_1} + \frac{\mu}{r_2}.$$

With this Lagrangian are associated the canonical moments

$$p_x = \frac{\delta L}{\delta x'} = x' - y,$$

$$p_y = \frac{\delta L}{\delta y'} = y' + x,$$

and the Hamiltonian

$$H = \frac{1}{2}(p_x^2 + p_y^2) + (yp_x - xp_y) - \left(\frac{1-\mu}{r_1} + \frac{\mu}{r_2}\right).$$

In this rotating frame the system of canonical equations takes the form

$$\frac{dx}{dt} = p_x - y, \qquad \frac{dp_x}{dt} = p_y - \frac{\delta V}{\delta x},$$

$$\frac{dy}{dt} = p_y - x, \qquad \frac{dp_y}{dt} = -p_x - \frac{\delta V}{\delta y}.$$

We can notice now that the Lagrangian and Hamiltonian in the rotating coordinate system are much more convenient for the calculations than these in the fixed coordinate system, for the reason that in the synodical frame the time t is not explicitly present in both functions. In the sidereal frame, both functions contain the time t in the distances r_1 and r_2, through the variable coordinates of m_1 and m_2.

In the preceding coordinate systems we always took the center of mass at the origin, but we shall use in the following developments two other rotating frames which are derived from the rotating barycentric system by a simple x translation. In one system (X, Y) the origin is at the largest mass m_1 and we shall call this system *geocentric*. In the other system, the *median* system, the origin is in the middle of both masses.

We define the *median* rotating coordinate system by the equations

$$x = X + x_0, \qquad y = Y, \qquad x_0 = \frac{1}{2} - \mu,$$

where x_0 is the distance between the center of mass and the middle of both masses m_1 and m_2. We can then obtain the Lagrangian and the Hamiltonian,

$$L = \frac{1}{2}(X'^2 + Y'^2) + (XY' - YX') + \frac{1}{2}(X^2 + Y^2)$$

$$+ x_0 X + \left(\frac{1-\mu}{r_1} + \frac{\mu}{r_2}\right),$$

$$H = \frac{1}{2}(p_X^2 + p_Y^2) + (p_X Y - p_Y X)$$

$$- \left(\frac{1-\mu}{r_1} + \frac{\mu}{r_2}\right) - x_0 X.$$

We can also define the *geocentric* rotating coordinate system (X, Y) by the simple translation

$$x = X - \mu, \qquad y = Y,$$

and the Hamiltonian becomes

$$H = \frac{1}{2}(p_X^2 + p_Y^2) + (p_X Y - p_Y X) - \left(\frac{1-\mu}{r_1} + \frac{\mu}{r_2}\right) + \mu X.$$

3. The Equations of Motion with Rotating Geocentric Polar Coordinates

We shall start this paragraph from the geocentric synodic coordinates (X, Y), and introduce a new set of coordinates by

$$X = r \cos\phi, \qquad Y = r \sin\phi.$$

The principal reason for interest in the geocentric polar coordinates is that they prepare a regularization of the geocentric singularity $r_1 = 0$. This regularization is performed by taking new polar coordinates (R, α) and then new Cartesian coordinates (ξ, η), by

$$r = R^2,$$

$$\phi = 2\alpha,$$

$$\xi = R \cos\alpha,$$

$$\eta = R \sin\alpha,$$

but the new (ξ, η) coordinates are not different from the parabolic coordinates, and we shall discuss them later.

It is possible to find the Hamiltonian in (r, ϕ) directly from the Hamiltonian in (X, Y) by a canonical extension. We can take the generating function

$$W(r, \phi, p_X, p_Y) = r \cos \phi \, p_X + r \sin \phi \, p_Y$$

and we define the moments by

$$p_r = \frac{\partial W}{\partial r} = \cos \phi \, p_X + \sin \phi \, p_Y,$$

$$p_\phi = \frac{\partial W}{\partial \phi} = -r \sin \phi \, p_X + r \cos \phi \, p_Y.$$

The inverse transformation can be obtained by

$$p_X = \frac{1}{r} (r \cos \phi \, p_r - \sin \phi \, p_\phi),$$

$$p_Y = \frac{1}{r} (r \sin \phi \, p_r + \cos \phi \, p_\phi),$$

as far as $r \neq 0$.

We have then the Hamiltonian in the simple form

$$H = \frac{1}{2} \left(p_r^2 + \frac{p_\phi^2}{r^2} \right) - p_\phi + \mu \cos \phi + V,$$

and the corresponding canonical equations of motion are

$$\frac{dr}{dt} = p_r, \qquad \frac{dp_r}{dt} = \frac{p_\phi^2}{r^3} - \frac{\partial V}{\partial r},$$

$$\frac{d\phi}{dt} = \frac{p_\phi}{r^2} - 1, \qquad \frac{dp_\phi}{dt} = -\frac{\partial V}{\partial \phi} + \mu \sin \phi.$$

4. The Equations of Motion with Biradial Coordinates

It is here our intention to give the equations of motion expressed with the variables (r_1, r_2). The variables (r_1, r_2) are related to the median synodical coordinates (X, Y) by

$$r_1^2 = \left(X + \frac{1}{2} \right)^2 + Y^2, \qquad 2X = (r_1^2 - r_2^2),$$

$$r_2^2 = \left(X - \frac{1}{2} \right)^2 + Y^2, \qquad 4Y^2 = 2S - 4X^2 - 1,$$

where

$$S = r_1^2 + r_2^2.$$

We have then

$$X' = r_1 r_1' - r_2 r_2',$$

$$2YY' = (1 - 2X) r_1 r_1' + (1 + 2X) r_2 r_2',$$

and we can express the whole Lagrangian in (r_1, r_2, r_1', r_2'). We find the result

$$L = \frac{A}{2} (r_1'^2 + r_2'^2) + B r_1' r_2' + \phi_1 r_1' + \phi_2 r_2'$$

$$+ \frac{1}{8} (2S - 1) + x_0 X - V,$$

where the four following auxiliary functions have been introduced:

$$A = \frac{r_1^2 r_2^2}{Y^2}, \qquad B = \frac{r_1 r_2}{2Y^2} (1 - S),$$

$$\phi_1 = \frac{1}{4Y} [2X - (2S - 1)] r_1,$$

$$\phi_2 = \frac{1}{4Y} [2X + (2S - 1)] r_2.$$

The canonical moments (p_1, p_2) associated with the variables (r_1, r_2) are then defined by the equations

$$p_1 = \frac{\delta L}{\delta r_1'} = A r_1' + B r_2' + \phi_1,$$

$$p_2 = \frac{\delta L}{\delta r_2'} = B r_1' + A r_2' + \phi_2,$$

and if the determinant $A^2 - B^2 = A$ is not zero, we can solve for r_1' and r_2':

$$r_1' = (p_1 - \phi_1) - \frac{B}{A} (p_2 - \phi_2),$$

$$r_2' = -\frac{B}{A} (p_1 - \phi_1) + (p_2 - \phi_2).$$

After a few simplifications we can then find the corresponding Hamiltonian:

$$H = \frac{1}{2} (p_1^2 + p_2^2) - \frac{1 - r_1^2 - r_2^2}{2 r_1 r_2} p_1 p_2$$

$$+ \frac{Y}{2} \left(\frac{p_1}{r_1} - \frac{p_2}{r_2} \right) + V + x_0 X.$$

The interest of the biradial coordinates is that they can be taken as a starting point for several new regularizing coordinate systems. For instance, the coordinate system (α, β) defined by the equations

$$r_1 = \alpha^2 + \cos^2 \beta, \qquad r_2 = \alpha^2 + \sin^2 \beta$$

removes simultaneously both singularities $r_1 = 0$, $r_2 = 0$, of the restricted three-body problem. But the (α, β) coordinates are related to coordinates which are described below.

5. The Equations of Motion with Elliptical Coordinates

The elliptical coordinates (r, s) can be defined from the biradial coordinates (r_1, r_2) by very simple formulas:

$$r = r_1 + r_2,$$
$$s = r_1 - r_2,$$
$$r_1 = \frac{1}{2}(r + s),$$
$$r_2 = \frac{1}{2}(r - s).$$

They are related to the median synodical rectangular coordinates by

$$2X = rs,$$
$$4Y^2 = (1 - s^2)(r^2 - 1).$$

We have then

$$X' = \frac{1}{2}(r's + rs'),$$

$$YY' = \frac{1}{4}[rr'(1 - s^2) - ss'(r^2 - 1)],$$

and the Lagrangian becomes

$$L = \frac{1}{8}(r^2 - s^2)\left[\frac{r'^2}{r^2 - 1} + \frac{s'^2}{1 - s^2}\right]$$
$$+ \frac{1}{4}\left[\left(\frac{1 - s^2}{r^2 - 1}\right)^{\frac{1}{2}} sr' - \left(\frac{r^2 - 1}{1 - s^2}\right)^{\frac{1}{2}} rs'\right]$$
$$+ \frac{1}{8}(r^2 + s^2 + 4x_0 rs) + 2\frac{r - 2x_0 s}{r^2 - s^2}.$$

The canonical moments (p_r, p_s) are defined by

$$p_r = \frac{1}{4}\frac{r^2 - s^2}{r^2 - 1} r' + \phi_r, \qquad p_s = \frac{1}{4}\frac{r^2 - s^2}{1 - s^2} s' + \phi_s,$$

where

$$\phi_r = \frac{s}{4}\left(\frac{1 - s^2}{r^2 - 1}\right)^{\frac{1}{2}}, \qquad \phi_s = -\frac{r}{4}\left(\frac{r^2 - 1}{1 - s^2}\right)^{\frac{1}{2}},$$

and the preceding equations can be solved for r', s':

$$r' = 4(p_r - \phi_r)\frac{r^2 - 1}{r^2 - s^2}, \qquad s' = 4(p_s - \phi_s)\frac{1 - s^2}{r^2 - s^2}.$$

We can then calculate the Hamiltonian, which becomes, after a few simplifications,

$$H = \frac{1}{r^2 - s^2}\{2(r^2 - 1)p_r^2 + 2(1 - s^2)p_s^2 - 2(r - 2x_0 s)$$
$$- \frac{1}{2}x_0 rs(r^2 - s^2) + [(1 - s^2)(r^2 - 1)]^{\frac{1}{2}}(rp_s - sp_r)\}.$$

The actual form of this Hamiltonian suggests that we can still write it in a more simple form by using the factor $1/(r^2 - s^2)$. As we have a Hamiltonian which does not contain explicitly the time t, we have the energy equation

$$H = \frac{1}{r^2 - s^2} H' = h,$$

where h is a constant. We can thus define a new Hamiltonian

$$\bar{H} = H' - (r^2 - s^2) h,$$

and we have then a new energy equation

$$\bar{H} = 0,$$

but this change of Hamiltonian corresponds also to a change of the independent variable t in τ by

$$dt = (r^2 - s^2) d\tau.$$

The complete new Hamiltonian can now be written:

$$\bar{H} = 2(r^2 - 1)p_r^2 + 2(1 - s^2)p_s^2$$
$$- 2(r - 2x_0 s) - \frac{1}{2}(r^2 - s^2)(x_0 rs + 2h)$$
$$+ [(1 - s^2)(r^2 - 1)]^{\frac{1}{2}}(rp_s - sp_r).$$

The corresponding canonical equations of motion are now:

$$\frac{dr}{d\tau} = 4(r^2 - 1)p_r - s[(1 - s^2)(r^2 - 1)]^{1/2},$$

$$\frac{ds}{d\tau} = 4(1 - s^2)p_s + r[(1 - s^2)(r^2 - 1)]^{1/2},$$

$$\frac{dp_r}{d\tau} = -4rp_r^2 + 2 + r(x_0 rs + 2h) + \frac{x_0}{2}(r^2 - s^2)s$$

$$- \frac{r(1 - s^2)(rp_s - sp_r)}{[(1 - s^2)(r^2 - 1)]^{1/2}} - [(1 - s^2)(r^2 - 1)]^{1/2} p_s,$$

$$\frac{dp_s}{d\tau} = +4sp_s^2 - 4x_0 - s(x_0 rs + 2h) + \frac{x_0}{2}(r^2 - s^2)r$$

$$+ \frac{s(r^2 - 1)(rp_s - sp_r)}{[(1 - s^2)(r^2 - 1)]^{1/2}} + [(1 - s^2)(r^2 - 1)]^{1/2} p_r.$$

Finally the elliptic coordinates (r, s) are related to Thiele's regularizing coordinates (E, F) by setting

$$r = \cosh F, \qquad s = \cos E,$$

but we shall describe these coordinates at the end of this text.

6. The Equations of Motion with the Regularizing Parabolic Coordinates (Ref. 1)

We can introduce the parabolic coordinates (ξ, η) from the barycentric synodic coordinates (x, y) by the equations

$$x = A + (\xi^2 - \eta^2), \qquad y = 2\xi\eta,$$

where the constant A takes the value $-\mu$, if we want to study the geocentric vicinity, and the value $1 - \mu$, if we want to study the selenocentric vicinity. When $A = -\mu$, the parabolic coordinates remove the singularity $r_1 = 0$, while for the other value, they remove the singularity $r_2 = 0$.

The equations of motion are then, in the Lagrangian form, for instance,

$$\ddot{\xi} - 8S\dot{\eta} = 8\xi\left\{\frac{3}{2}S^2 + A(S + D) + \frac{A^2}{2}\right.$$

$$\left. + E_0 - A\frac{S - 2D + B}{[(S^2 - 2BD + 1)^3]^{1/2}}\right\},$$

$$\ddot{\eta} + 8S\dot{\xi} = 8\eta\left\{\frac{3}{2}S^2 - A(S - D) + \frac{A^2}{2}\right.$$

$$\left. + E_0 - A\frac{(-S - 2D + B)}{[(S^2 - 2BD + 1)^3]^{1/2}}\right\},$$

where E_0 is the initial energy constant, and where B takes the value $+1$ in the geocentric case and -1 in the selenocentric case. We have introduced here the functions

$$S = \xi^2 + \eta^2, \qquad D = \xi^2 - \eta^2.$$

The dot represents a derivative with respect to a new "regularizing time" s, related to the physical time t by the quadrature

$$dt = 4S \, ds.$$

7. The Regularized Equations of Motion with the Generalized Birkhoff Transformation

The most convenient way to introduce Birkhoff's generalized coordinates (ξ, η) is to start from the median synodic coordinates (X, Y) and to use complex variables defined by

$$Z = X + iY, \qquad \zeta = \xi + i\eta.$$

Then the coordinate transformation is represented by a conformal mapping between the two complex planes Z and ζ,

$$Z = Z(\zeta).$$

The Hamiltonian of the restricted three-body problem with the new variables (ξ, η) and their corresponding canonical moments (p_ξ, p_η) takes the general form

$$H = \left|\frac{dZ}{d\zeta}\right|^{-2}\left[\frac{1}{2}(p_\xi^2 + p_\eta^2) - (A_\xi p_\xi + A_\eta p_\eta)\right.$$

$$\left. + \left|\frac{dZ}{d\zeta}\right|^2 \tilde{V}(\xi, \eta)\right],$$

where

$$A_\xi = -\frac{1}{2}\frac{\partial}{\partial\eta}(X^2 + Y^2),$$

$$A_\eta = +\frac{1}{2}\frac{\partial}{\partial\xi}(X^2 + Y^2),$$

$$\tilde{V} = V - x_0 X,$$

where $X^2 + Y^2$ is a function of (ξ, η).

If we introduce now a new independent variable s, related with the time variable t by the relation

$$dt = \left|\frac{dz}{d\zeta}\right|^2 ds,$$

we obtain a new Hamiltonian

$$\overline{H} = \frac{1}{2}(p_\xi^2 + p_\eta^2) - (A_\xi p_\xi + A_\eta p_\eta)$$
$$+ \left|\frac{dz}{d\zeta}\right|^2 [\tilde{V}(\xi, \eta) - h],$$

where h is the energy constant corresponding to the energy integral $H = h$.

The Lagrangian equations of motion may then be written with the new variables ξ, η, s, in the form

$$\ddot{\xi} - 2\left|\frac{dz}{d\zeta}\right|^2 \dot{\eta} = -\frac{\partial \overline{V}}{\partial \xi}, \qquad \ddot{\eta} + 2\left|\frac{dz}{d\zeta}\right|^2 \dot{\xi} = -\frac{\partial \overline{V}}{\partial \eta},$$

where

$$\overline{V} = \left|\frac{dz}{d\zeta}\right|^2 [\tilde{V}(\xi, \eta) - h] = \left|\frac{dz}{d\zeta}\right|^2 (V - x_0 X - h).$$

We also make here the convention that a prime always indicates a derivative with respect to the physical time t, while a dot indicates a derivative with respect to the new regularized time s.

In the generalized Birkhoff transformation the conformal mapping which is used is

$$Z = \frac{1}{4}\left(\zeta^n + \frac{1}{\zeta^n}\right).$$

We suppose here that the parameter n is an integer number, different from zero, although this parameter could be real. We have here a mapping which has two critical points, $Z = \pm\frac{1}{2}$, corresponding with the two singularities of the three-body problem. These two critical points are at the same time branch points.

We can also write the transformation in a real form

$$X = \frac{1}{4S^n}(S^n + 1)X_1, \qquad Y = \frac{1}{4S^n}(S^n - 1)Y_1,$$

where

$$S^n = (X_1^2 + Y_1^2) = (\xi^2 + \eta^2)^n = |\zeta|^{2n},$$

and where X_1 and Y_1 are both homogeneous polynomials of degree n in ξ and η. They may be expressed in terms of the binomial coefficients C_n^i,

$$X_1 = \xi^n - C_n^2 \eta^2 \xi^{n-2} + C_n^4 \eta^4 \xi^{n-4} - C_n^6 \eta^6 \xi^{n-6} + \cdots,$$
$$Y_1 = C_n^1 \eta \xi^{n-1} - C_n^3 \eta^3 \xi^{n-3} + C_n^5 \eta^5 \xi^{n-5} - \cdots.$$

For instance, when $n = 1$, we have

$$X_1 = \xi_1, \qquad Y_1 = \eta,$$

and when $n = 2$,

$$X_1 = \xi^2 - \eta^2, \qquad Y_1 = 2\xi\eta.$$

We can now write r_1, r_2, as functions of ξ, η,

$$r_1 = \frac{1}{4(S^n)^{1/2}} [(S^n + 1) - 2X_1],$$

$$r_2 = \frac{1}{4(S^n)^{1/2}} [(S^n + 1) + 2X_1],$$

and the Jacobian of the transformation can be written in the form

$$J = \left|\frac{dz}{d\zeta}\right|^2 = \frac{r_1 r_2 n^2}{S} = \frac{n^2}{16 S^{n+1}}[(S^n + 1)^2 - 4X_1^2].$$

With the above results we can now write the Hamiltonian of the restricted three-body problem in terms of the new variables (ξ, η). We have the Hamiltonian and the energy equation

$$H = \frac{1}{r_1 r_2}\left\{\frac{S}{2n^2}(p_\xi^2 + p_\eta^2) - \frac{S}{n^2}(A_\xi p_\xi + A_\eta p_\eta)\right.$$
$$\left. - [(1 - \mu)r_1 + \mu r_2] - x_0 X r_1 r_2\right\} = h.$$

By the introduction of the new time variable s,

$$dt = r_1 r_2 ds,$$

we obtain the new Hamiltonian,

$$\overline{H} = \frac{S}{2n^2}(p_\xi^2 + p_\eta^2) - \frac{S}{n^2}(A_\xi p_\xi + A_\eta p_\eta)$$
$$- [(1 - \mu)r_1 + \mu r_2] - (x_0 X + h)r_1 r_2.$$

It is now elementary to derive the canonical equations from the above Hamiltonian. All the partial derivatives

that are needed have been given in Ref. 2, which contains a detailed study of this regularizing coordinate transformation.

When $n = 1$, we have Birkhoff's well known regularization and when $n = 2$, we have a regularization corresponding with works done by G. Lemaitre (Ref. 3) and also by R. F. Arenstorf (Ref. 4).

8. The Equations of Motion with Thiele's Coordinates (Ref. 5)

The coordinate transformation $z = \frac{1}{4}(\zeta^n + \zeta^{-n})$ takes an interesting form if we introduce a new transformation, with the complex notations,

$$\theta = e^{i\theta},$$

or in the real form

$$\xi = e^{-F} \cos E, \qquad E = \arctan \frac{\eta}{\xi},$$

$$\eta = e^{-F} \sin E, \qquad F = -\frac{1}{2} \log_e (\xi^2 + \eta^2),$$

where θ is a complex variable

$$\theta = E + iF.$$

Thus we see that (E, F) are a special system of polar coordinates in the (ξ, η) plane; e^{-F} is a radius, while E is an angle.

With the introduction of the θ variable, the Z-ζ correspondence becomes now a Z-θ correspondence which can be written in the form

$$Z = \frac{1}{2} \cos n\theta.$$

The real form of this conformal mapping is very easy to obtain, and is not restricted to the integer values of n, because we have

$$\zeta^n = e^{-nF} (\cos nE + i \sin nE),$$

$$X_1 = e^{-nF} \cos nE, \qquad Y_1 = e^{-nF} \sin nE,$$

with

$$Y_1^2 + Y_1^2 = S^n = e^{-2nF}.$$

The real form of our mapping is thus

$$X = \frac{1}{2} \cos nE \cosh nF,$$

$$Y = -\frac{1}{2} \sin nE \sinh nF.$$

The Hamiltonian for the restricted three-body problem can now be written in the form

$$H = \frac{1}{n^2 r_1 r_2} [2(p_E^2 + p_F^2)$$

$$- (p_E \sinh nF \cosh nF + p_F \sin nE \cos nE)$$

$$+ 2(2x_0 \cos nE - \cosh nF) - 2x_0 r_1 r_2 \cos nE \cosh nF].$$

Thiele's coordinates correspond to the value $n = 1$. A detailed study of the particular case $n = 1$ has been given in Ref. 6.

B. On the Second Variation of an Optimal Trajectory

C. G. Pfeiffer

1. Introduction

Recent work in the calculus of variations has been directed toward the optimal final-value control problem, which can be described as the task of controlling the trajectory of a point mass so as to achieve a desired final state while at the same time minimizing (or maximizing) a given function of the final state variables (Ref. 7 and footnote[1]). For example, this problem arises when it is desired to construct a steering law for a rocket vehicle which is to inject a spacecraft into satellite orbit with maximum horizontal speed, where the burnout altitude, the burnout flight path angle, and the duration of thrusting are specified (Ref. 8).

[1]Melbourne, W. G., Unpublished Lecture Notes on the Calculus of Variations, Jet Propulsion Laboratory, Pasadena, California, 1963.

The system analyst is usually interested, first of all, in describing the mathematical properties of the optimal control function, and, secondly, in developing both analytical and numerical techniques for constructing such a function. It is the first problem that is the concern of this paper, that is, we shall seek necessary and sufficient conditions for a trajectory to yield a minimum value of the function to be optimized, subject to the imposed boundary conditions.

The problem discussed here will be somewhat simplified in order to facilitate the exposition. Thus, it will be assumed that there is only one boundary condition to meet, that the final time is fixed, that there is only one control variable to consider, that the control variable is not bounded, and that the trajectory is not a singular extremal. (A more general case is discussed in Ref. 9.) In the perturbation analysis we shall limit ourselves to considering only small deviations in the control variable. By this restriction we are dealing with what is called a "weak" variation in the terminology of the classical calculus of variations.

The notation used is as follows: The independent variable is t, which may be thought of as time; T is the (fixed) final time; other capital letters are matrices; I is the identity matrix; column vectors are denoted by a bar ($-$) over a small letter; the transpose of a vector or matrix is indicated by the superscript $'$; δ refers to the variation of the indicated quantity from its (known) nominal value (thus $\delta \bar{x}(t) = \bar{x}(t) - \bar{x}(t)_{nominal}$). The notation (t) will occasionally be omitted in equations in order to simplify the notation.

2. Formulation of the Problem

Let the rate of change of the state of the system be given by

$$\frac{d}{dt}\bar{x} = \bar{f}(\bar{x}, y, t) \qquad (1)$$

where

$$\bar{x}' = (x_1, x_2, \cdots, x_n)$$

is the state vector of the system (such as the position and velocity coordinates of a rocket vehicle); $y(t)$ is the continuous control function which is to be determined (such as the steering angle of the rocket vehicle); and

$$\bar{f}' = [f_1(\bar{x}, y, t), f_2(\bar{x}, y, t), \cdots, f_n(\bar{x}, y, t)]$$

is a known vector function of $\bar{x}$ and y, and is differentiable up to second order with respect to these quantities. Given some initial condition $\bar{x}(0)$, the problem is to choose the control variable $y(t)$ in such a way as to minimize at some fixed final time T the function $\beta_1(\bar{x})$, subject to the constraint $\beta_2(\bar{x}) = 0$. The $\beta_i(\bar{x})$ will be called the "boundary functions." Without loss of generality it can be assumed that the $\beta_i(\bar{x})$ are linear functions of the final state $\bar{x}$. Employing the Lagrange multiplier technique, we seek to minimize the performance index function

$$p(\bar{x}) = \beta_1(\bar{x}) + v\beta_2(\bar{x}) \qquad (2)$$

where v is the constant Lagrange multiplier. (The v is usually obtained by a search procedure in the construction of a nominal trajectory.) This is the well known Mayer formulation of the optimization problem in the calculus of variations, which includes the case of minimizing a functional of $\bar{x}(t)$ and $y(t)$ and/or of applying integral constraints.

3. The First Variation of the Boundary Functions

We suppose that a nominal trajectory (not necessarily optimal) exists, and define the variations of the state and control variables to be

$$\delta \bar{x}(t) = \bar{x}(t) - \bar{x}(t)_{nominal}$$

$$\delta y(t) = y(t) - y(t)_{nominal}$$

Seeking an expression for the first variation of the performance index function p, we construct a Taylor series expansion of Eq. (1) in terms of $\delta \bar{x}$ and δy. Thus,

$$\frac{d}{dt}\delta \bar{x}(t) = F(t)\,\delta \bar{x}(t) + G(t)\,\delta y(t)$$

$$+ \text{ terms of the type } [\delta y^2(t)]$$

$$+ \text{ terms of the type } [\delta x_i(t)\,\delta y(t)]$$

$$+ \text{ terms of the type } [\delta x_i(t)\,\delta x_j(t)]$$

$$+ \text{ higher order terms} \qquad (3)$$

where $F(t)$ is an n by n matrix with elements

$$f_{ij} = \left(\frac{\partial f_i}{\partial x_j}\right)$$

and $G(t)$ is an n by 1 matrix with elements

$$g_{i1} = \left(\frac{\partial f_i}{\partial y}\right)$$

We introduce the state transition matrix $U(\tau, t)$, which is an n by n matrix with elements interpreted as

$$u_{ij}(\tau, t) = \left(\frac{\partial x_i(\tau)}{\partial x_j(t)}\right)$$

We compute

$$\frac{d}{dt}[U\delta\bar{x}] = \left[\frac{d}{dt}U\right]\delta\bar{x} + U\left[\frac{d}{dt}\delta\bar{x}\right] \qquad (4)$$

and apply Eq. (3) and well known properties of $U(\tau, t)$ to obtain

$$\delta\bar{x}(\tau) = U(\tau, t)\delta\bar{x}(t) + \int_t^\tau U(\tau, s)G(s)\delta y(s)\,ds$$

$$+ \text{ higher order terms} \qquad (5)$$

The variation of the boundary function $\beta_i(\bar{x})$ defined above is

$$\delta\beta_i = \overline{\nabla}'\beta_i\delta\bar{x}(T) = \sum_{j=1}^n \left(\frac{\partial\beta_i}{\partial x_j}\right)\delta x_j(T)$$

where $\overline{\nabla}\beta_i$ is the gradient of β_i. Applying Eq. (5) with $\tau = T$, we have

$$\delta\beta_i = \bar{\lambda}_i'(t)\delta\bar{x}(t) + \int_t^T \eta_i(s)\delta y(s)\,ds$$

$$+ \text{ higher order terms} \qquad (6)$$

where

$$\bar{\lambda}_i'(t) \triangleq \overline{\nabla}'\beta_i U(T, t)$$

$$\eta(t) \triangleq \bar{\lambda}_i'(t)G(t)$$

The first two terms on the right hand side of Eq. (6) express the first variation of $\delta\beta_i$ in the desired functional form. The elements of $\bar{\lambda}_i(t)$ are called the differential corrections for the function β_i, and $\eta_i(t)$ is the influence (Green's) function.

4. The First and Second Necessary Conditions for Optimality

Consider the performance index function p defined by Eq. (2), and, as in the previous section, construct the differential correction vector $\bar{\lambda}_p$ and the influence function η_p. On an optimal trajectory it is necessary that $\eta_p(t) = 0$, because, with no bounds on $y(t)$, the first variation of p must be stationary with respect to the control. Thus, we have:

(1) *The first necessary condition.* A necessary condition for a trajectory to be optimal (stationary) with respect to the (unbounded) control is that for all $0 \le t \le T$,

$$\eta_p(t) = \eta_1(t) + \nu\eta_2(t) = 0$$

This condition also follows from the well known Pontryagin principle (Ref. 7) where the generalized Hamiltonian is defined as

$$h \triangleq \bar{\lambda}_p'(t)\bar{f}(\bar{x}, y, t)$$

The Pontryagin principle states that h must be a minimum with respect to the control if the trajectory is minimizing. Since we have assumed that $\bar{f}$ is differentiable with respect to y, and since y is unbounded, we have $(\partial h/\partial y) = \eta_p = 0$.

(2) *The second necessary condition.* Implicit in the Pontryagin principle is a further necessary condition, which is

$$\frac{\partial^2 h}{\partial y^2} \ge 0 \qquad \text{for all } 0 \le t \le T$$

This result can also be obtained from a consideration of the second variation of the penalty functional with respect to the control, which will be developed below. For the purpose of this discussion it will be assumed that the inequality holds strictly, which is to say that singular extremals $[(\partial^2 h/\partial y^2)(t) = 0]$ are excluded from consideration.

5. The Second Variation of the Function to be Minimized

The second variation of the function $\beta_1(\bar{x})$ is obtained by considering only the contributions of second order

terms in Eq. (6); thus,

$$\delta^2\beta_1 \triangleq \int_0^T [\text{terms of the type } \delta y^2(t)] \, dt$$

$$+ \int_0^T [\text{terms of the type } \delta x_i(t) \, \delta y(t)] \, dt$$

$$+ \int_0^T [\text{terms of the type } \delta x_i(t) \, \delta x_j(t)] \, dt \qquad (7)$$

Let us define

$$W(\tau, t) \triangleq \begin{cases} U(\tau, t) \, G(t) & \text{for } \tau \geq t \\ 0 & \text{for } \tau < t \end{cases}$$

and express the quantities $\delta x_i(t)$ in Eq. (7) as

$$\delta \bar{x}(\tau) = U(\tau, t) \, \delta \bar{x}(t) + \int_t^T W(\tau, s) \, \delta y(s) \, ds$$

Thus we have

$$\delta^2\beta_1 = \int_0^T a(t) \, \delta y^2(t) \, dt$$

$$+ \int_0^T \int_0^T b(\tau, t) \, \delta y(\tau) \, \delta y(t) \, d\tau \, dt$$

$$+ \int_0^T \int_0^T \int_0^T c(\tau, t, s) \, \delta y(\tau) \, \delta y(t) \, d\tau \, dt \, ds$$

where $a(t)$, $b(\tau, t)$, and $c(\tau, t, s)$ are kernels obtained from the above described analysis. Suppose we assume[2] $a(t) > 0$ for $0 \leq t \leq T$, and define

$$u(t) \triangleq [a(t)]^{1/2} \, \delta y(t)$$

Then, after reducing the triple integral by integrating once with respect to s, $\delta^2\beta_1$ can be put into the symmetric form

$$\delta^2\beta_1 = \int_0^T u^2(t) \, dt + \int_0^T \int_0^T k(t, \tau) \, u(t) \, u(\tau) \, dt \, d\tau \qquad (8)$$

where $k(t, \tau)$ is a symmetric kernel. This is the desired second order functional expansion of $\beta_1(\bar{x})$.

_____

[2]This assumption is suggested by the second necessary condition, but is not guaranteed unless $\nu = 0$. The case where it does not hold will not be considered here.

6. Application of the Hilbert-Schmidt Theory of Integral Equations

The symmetric form of the second variation developed above allows us to apply the Hilbert-Schmidt theory of integral equations. It is shown in Chapter V of Ref. 10 that to every real continuous, symmetric kernel $k(t, \tau)$, which is not identically zero over the interval $(0, T)$, there belongs a set of real "characteristic constants"

$$(\cdots \leq \omega_1 \leq \omega_2 \leq \cdots \leq \omega_i \leq \omega_{i+1} \leq \cdots)$$

and with each characteristic constant there is associated a real "fundamental function" $\phi_i(\tau)$, such that the set $\{\phi_i(\tau)\}$ is a complete normalized orthogonal system with the properties

$$\phi_i(t) = \omega_i \int_0^T k(t, \tau) \, \phi_i(\tau) \, d\tau$$

$$\int_0^T \phi_i(\tau) \, \phi_j(\tau) \, d\tau = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

It follows that if a function $v(t)$ can be represented in the form

$$v(t) = \int_0^T k(t, \tau) \, u(\tau) \, d\tau$$

where $u(\tau)$ is continuous on $(0, T)$, then

$$v(t) = \sum_i \frac{c_i}{\omega_i} \, \phi_i(t)$$

where

$$c_i = \int_0^T u(\tau) \, \phi_i(\tau) \, d\tau.$$

We therefore conclude that Eq. (8) can be written

$$\delta^2\beta_1 = \int_0^T u^2(t) \, dt + \sum_i \frac{c_i^2}{\omega_i} \qquad (9)$$

This result will be employed to prove the following theorem.

7. The Second Variation Theorem

Suppose that the first and second necessary conditions are satisfied on a given nominal trajectory, and let us consider weak variations in the control variable as described above. We then conclude that a further necessary

condition for the trajectory to be minimizing is that $\alpha_i \geqq 0$ for all i, where α_i is a solution of

$$0 = \int_0^T g^2 \, dt - \sum_i \left[\int_0^T g \, \phi_i \, dt \right]^2 [(1 - \alpha)\,\omega_i + 1]^{-1}$$

and $g(t) = [\eta_2(t)]\,[a(t)]^{-\frac{1}{2}}$. A sufficient condition is that $\alpha_i > 0$.

Proof. Since we are considering only weak (small) variations of the control variable, the first and second variations dominate the higher order terms in the expanded form of the boundary functions. From the first necessary condition it follows that the first variation of β_1 is zero if the first variation of β_2 is zero. Our proof therefore consists of finding the extreme (in particular, the minimal) values of $\delta^2\beta_1$ for non-zero $u(t)$, subject to the constraint

$$0 = \int_0^T \eta_2(t)\,\delta y(t)\,dt = \int_0^T g(t)\,u(t)\,dt \qquad (10)$$

Clearly $(\delta^2\beta_1)_{min} \geqq 0$ is a necessary condition, $(\delta^2\beta_1)_{min} > 0$ is a sufficient condition, and if $(\delta^2\beta_1)_{min} = 0$ no conclusion about the minimal nature of the trajectory can be reached from considering the second variation. To restrict ourselves from considering the trivial solution $u(t) = 0$, we shall set

$$\int_0^T u^2(t)\,dt = \epsilon > 0 \qquad (11)$$

where ϵ is a small number. Thus we seek to extremalize

$$q = \delta^2\beta_1 + \alpha \left[\epsilon - \int_0^T u^2 \, dt \right] + 2\sigma \int_0^T g\,u\,dt$$

where α and σ are constant Lagrange multipliers. This is analogous to the classical accessory minimum problem (Ref. 11). It follows from the variational calculus that the extremalizing $u(t)$ is characterized by

$$\alpha u(t) = u(t) + \int_0^T k(t,\tau)\,u(\tau)\,d\tau + \sigma g(t)$$

$$= u(t) + \sum_i \frac{c_i \phi_i(t)}{\omega_i} + \sigma g(t) \qquad (12)$$

Multiplying Eq. (12) by $u(t)$ and integrating, and applying Eqs. (9), (10), and (11), we have

$$\alpha \epsilon = \int_0^T u^2 \, dt + \sum_i \frac{c_i^2}{\omega_i} = (\delta^2\beta_1)_{extremum}$$

Thus $\alpha \geqq 0$ is a necessary condition, $\alpha > 0$ is a sufficient condition, and we proceed to find an expression for α.

We multiply Eq. (12) by $\phi_i(t)$ and integrate to obtain

$$c_i \left[1 + \frac{1}{\omega_i} - \alpha \right] + \sigma \int_0^T g(t)\,\phi_i(t)\,dt = 0$$

We multiply Eq. (12) by $g(t)$ and integrate to obtain

$$\sum_i \left(\frac{c_i}{\omega_i} \right) \left[\int_0^T \phi_i\,g\,dt \right] + \sigma \int_0^T g^2 \, dt = 0$$

Combining these last two equations we have

$$\sigma \left\{ \int_0^T g^2 \, dt - \sum_i \left[\int_0^T \phi_i g \, dt \right]^2 [(1-\alpha)\,\omega_i + 1]^{-1} \right\} = 0$$

Since this equation must hold for non-zero σ, the theorem is proved. (The existence of at least one solution α is assured, since there must be at least one extreme value of $\delta^2\beta_1$.)

8. Discussion

We have summarized here a technique for analyzing the second variation of the boundary functions on an optimal trajectory (Ref. 9), and we have applied the Hilbert-Schmidt theory of integral equations to develop the second variation theorem. The necessary and sufficient conditions obtained must be analogous to similar results obtained for the Bolza problem, as discussed in Ref. 11. The Hilbert-Schmidt theory of integral equations provides a straightforward and elegant general treatment of the optimal control problem, and evidently has been similarly employed by other authors (see "A Bibliography for the Problem of Bolza" in Ref. 11). The second order functional expansion developed here will probably have other important applications in control theory.

C. A Probability Distribution Associated with a Multiple Midcourse Maneuver

C. B. Solloway

In a simplified version of the multiple midcourse error analysis the following problem arises: Let x be a normal random variable with mean μ_x and standard deviation σ_x. Let y be a random variable statistically independent of x, which with probability p takes the value α and with probability $q = 1 - p$ has a normal type distribution with mean β and standard deviation τ. Let $z = x + y$. Determine the density and distribution functions of z.

For instance, considering a one-dimensional multiple midcourse maneuver with two midcourse corrections available, x would represent the error in the first midcourse correction and y the error in the second. In this case, generally μ_x, α and β would be zero and p would represent the probability that a second correction was not needed; z would represent the error due to the complete sequence of correction maneuvers.

The following results are easily obtained:

$$p_1(x) = \frac{1}{(2\pi)^{1/2}\sigma_x} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]$$
$$-\infty < x < \infty \qquad (1)$$

$$p_2(y) = p\delta(y - \alpha)$$
$$+ \frac{q}{(2\pi)^{1/2}\tau} \exp\left[-\frac{1}{2}\left(\frac{y-\beta}{\tau}\right)^2\right]$$
$$-\infty < y < \infty \qquad (2)$$

$$E[y] = \mu_y = p\alpha + q\beta$$
$$V[y] = \sigma_y^2 = q\tau^2 + pq(\alpha - \beta)^2 \qquad (3)$$

$$p(z) = \frac{p}{(2\pi)^{1/2}\sigma_x} \exp\left[-\frac{(z-\mu_x-\alpha)^2}{2\sigma_x^2}\right]$$
$$+ \frac{q}{(2\pi)^{1/2}(\sigma_x^2+\tau^2)^{1/2}} \exp\left[-\frac{(z-\mu_x-\beta)^2}{2(\sigma_x^2+\tau^2)}\right]$$
$$-\infty < z < \infty \qquad (4)$$

$$E[z] = \mu_z = \mu_x + p\alpha + q\beta$$
$$V[z] = \sigma_z^2 = \sigma_x^2 + q\tau^2 + pq(\alpha - \beta)^2 \qquad (5)$$

$$Pr\{z > k\} = \int_k^\infty p(z)\,dz$$
$$= 0.5 - p\phi\left(\frac{k-\mu_x-\alpha}{\sigma_x}\right) - q\phi\left[\frac{k-\mu_x-\beta}{(\sigma_x^2+\tau^2)^{1/2}}\right] \qquad (6)$$

where

$$\phi(x) = \int_0^x \frac{1}{(2\pi)^{1/2}} \exp(-t^2/2)\,dt \qquad (7)$$

which is the most commonly tabulated indefinite integral of the standard normal distribution.

Another quantity of interest is the central deviation

$$Pr\{|z-\mu_z| < k\sigma_z\} = \int_{\mu_z - k\sigma_z}^{\mu_z + k\sigma_z} p(z)\,dz$$
$$= p\left|\phi\left(\left|\frac{q(\beta-\alpha)+k\sigma_z}{\sigma_x}\right|\right)\right.$$
$$- \text{sgn}\,[q^2(\beta-\alpha)^2 - k^2\sigma_z^2]$$
$$\left.\times \phi\left(\left|\frac{q(\beta-\alpha)-k\sigma_z}{\sigma_x}\right|\right)\right|$$
$$+ q\left|\phi\left(\left|\frac{p(\alpha-\beta)+k\sigma_z}{(\sigma_x^2+\tau^2)^{1/2}}\right|\right)\right.$$
$$- \text{sgn}\,[p^2(\alpha-\beta)^2 - k^2\sigma_z^2]$$
$$\left.\times \phi\left(\left|\frac{p(\alpha-\beta)-k\sigma_z}{(\sigma_x^2+\tau^2)^{1/2}}\right|\right)\right| \qquad (8)$$

where

$$\text{sgn}\,x = 1, \qquad x > 0$$
$$\text{sgn}\,x = -1, \qquad x < 0$$

These formulas simplify somewhat if $\mu_x = \alpha = \beta = 0$. We then have

$$E[z] = 0$$
$$V[z] = \sigma_x^2 + q\tau^2 \qquad (9)$$

$$Pr\{z > k\} = 0.5 - p\phi\left(\frac{k}{\sigma_x}\right) - q\phi\left[\frac{k}{(\sigma_x^2+\tau^2)^{1/2}}\right] \qquad (10)$$

$$Pr\{|z| \leq k\sigma_z\} = 2p\phi\left(\frac{k\sigma_z}{\sigma_x}\right) + 2q\phi\left[\frac{k\sigma_z}{(\sigma_x^2+\tau^2)^{1/2}}\right] \qquad (11)$$

If further investigations prove it necessary, it will be easy to generate tables for these functions.

The results are immediately generalized to the multidimensional case. Thus, if

$$p_1(x) = \frac{1}{(2\pi)^{n/2}|\Lambda_x|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - \mu_x)^T \Lambda_x^{-1}(x - \mu_x)\right]$$

(12)

$$p_2(y) = p\delta(y - \alpha)$$
$$+ \frac{q}{(2\pi)^{n/2}|\Lambda_\tau|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(y - \beta)^T \Lambda_\tau^{-1}(y - \beta)\right]$$

(13)

where x, μ_x, y, α, and β are n vectors and Λ_x and Λ_τ are non-singular covariance matrices, then, if $z = x + y$, we have

$$p(z) = \frac{p}{(2\pi)^{n/2}|\Lambda_x|^{\frac{1}{2}}}$$
$$\times \exp\left[-\frac{1}{2}(z - \mu_x - \alpha)^T \Lambda_x^{-1}(z - \mu_x - \alpha)\right]$$
$$+ \frac{q}{(2\pi)^{n/2}|\Lambda_x + \Lambda_\tau|^{\frac{1}{2}}}$$
$$\times \exp\left[-\frac{1}{2}(z - \mu_x - \beta)^T (\Lambda_x + \Lambda_\tau)^{-1}(z - \mu_x - \beta)\right]$$

(14)

and

$$E[x] = \mu_x; \; E[(x - \mu_x)(x - \mu_x)^T] = \Lambda_x$$

$$E[y] = p\alpha + q\beta; \quad E\{[y - E(y)][y - E(y)]^T\}$$
$$= q\Lambda_\tau + pq(\alpha - \beta)(\alpha - \beta)^T$$

$$E[z] = \mu_x + p\alpha + q\beta; \quad E\{[z - E(z)][z - E(z)]^T\}$$
$$= \Lambda_x + q\Lambda_\tau + pq(\alpha - \beta)(\alpha - \beta)^T$$
$$= \Lambda_z$$

(15)

As before, we can still ask, e.g., $Pr\{z^T\Lambda_z^{-1}z \leq k^2\}$ (which is the analogue of Eq. (11) if $\mu_x = \alpha = \beta = 0$) but now the integrals are quite intractable. (The problem is that of integrating a *circular* multidimensional Gaussian distribution over an *ellipsoidal* region. If necessary, tables can be prepared here as well.)

D. A Launch Probability Study —Surveyor

C. B. Solloway

The following problems arose in connection with the launch probabilities associated with the *Surveyor* vehicle:

(1) What is the probability of having exactly k successful launches (countdowns and firings) in n days under the following conditions:

   (a) Only one trial per day is permitted. If it fails, the trial is repeated the next day. If it succeeds, the next trial is postponed one day (turnaround time).

   (b) Each trial is independent; the probability of success on any trial is p.

   (c) There are exactly k vehicles available.

(2) What is the probability in the previous problem if the third condition is altered so that there are at least $(k + 1)$ vehicles available?

The answers to the two problems are easily found to be the following: For Problem (1), let $P_{n,k}$ equal the probability of obtaining the kth success on the nth day. Then

$$P_{n,k} = \binom{n - k}{k - 1} p^k q^{n-2k+1} \qquad 2k \leq n + 1$$
$$= 0 \qquad\qquad\qquad \text{otherwise}$$

(1)

where $q = 1 - p$ and

$$\binom{r}{s} = \frac{r!}{s!(r - s)!}$$

is the binomial coefficient. In particular,

$P_{n,0} = q^n$	(not given by Eq. 1)	(2)
$P_{n,1} = pq^{n-1}$	$n \geq 1$	(3)
$P_{n,2} = (n - 2)p^2q^{n-3}$	$n \geq 3$	(4)
$P_{n,3} = \frac{1}{2}(n - 3)(n - 4)p^3q^{n-5}$	$n \geq 5$	(5)

For example,

$$P_{3,2} = p^2 \qquad [S - S]$$
$$P_{4,2} = 2p^2q \qquad [S - FS \text{ or } FS - S]$$

For Problem (2), let $P_{n,k}$ equal the probability of having exactly k successes in n trials. Then

$$P_{n,k} = \binom{n-k}{k} p^k q^{n-2k} + \binom{n-k}{k-1} p^k q^{n-2k+1} \qquad (6)$$

where the standard convention is adopted that

$$\binom{r}{s} = 0$$

if $s > r > 0$ or $r > 0 > s$, r, s integers. In particular,

$$P_{n,0} = q^n \qquad\qquad n \geq 0 \qquad (7)$$

$$P_{n,1} = (n-1) pq^{n-2} + pq^{n-1} \qquad n \geq 1 \qquad (8)$$

$$P_{n,2} = \frac{1}{2}(n-2)(n-3) p^2 q^{n-4} + (n-2) p^2 q^{n-3}$$
$$n \geq 3 \qquad (9)$$

For example,

$P_{1,1} = p$	[S]
$P_{2,1} = p + pq$	[S— or FS]
$P_{3,2} = p^2$	[S — S]
$P_{4,2} = p^2 + 2p^2q$	[S — S— or FS — S or S — FS]

The analysis is given in the following paragraphs.

For Problem (1) the kth success must occur on the nth day and no more trials are available since all vehicles have been expended. This leaves $(k-1)$ prior successes which occupy $2(k-1)$ days and hence, leave

$$(n-1) - 2(k-1) = n - 2k + 1$$

failures to be accounted for. The number of distinct ways these

$$(k-1) + (n-2k+1) = n - k$$

objects can be permuted is

$$\binom{n-k}{k-1}$$

and the probability associated with each permutation is $p^{k-1} q^{n-2k+1}$. Hence, the result.

In Problem (2), the kth success can either occur on or before the nth day. If it occurs on the nth day, the analysis and answer are given by Problem (1). If it occurs prior to the nth day the k successes occupy $2k$ days and hence, leave $n - 2k$ days on which failures occurred. The number of distinct ways these $k + (n - 2k) = n - k$ objects can be permuted is

$$\binom{n-k}{k}$$

and the probability associated with each permutation is $p^k q^{n-k}$. The sum of the two probabilities yields the desired result.

It is interesting to solve the latter problem by the method of difference equations. $P_{n,k}$ satisfies the linear, homogeneous partial difference equation

$$P_{n,k} = qP_{n-1,k} + pP_{n-2,k-1} \qquad \begin{array}{l} n = 2,3,4 \cdots \\ k = 1,2,3 \cdots \end{array} \qquad (10)$$

which is deduced as follows: a failure is experienced the first day and then k successes are obtained in $(n-1)$ days or a success is obtained the first day, there is a day's delay, and then $(k-1)$ successes are obtained in $(n-2)$ days. This difference equation is subject to the initial conditions

$$P_{n,0} = q^n \qquad\qquad n = 0,1,2, \cdots$$

$$P_{0,k} = \begin{cases} 1 & k = 0 \\ 0 & k = 1,2,3, \cdots \end{cases}$$
$$\qquad (11)$$

$$P_{1,k} = \begin{cases} q & k = 0 \\ p & k = 1 \\ 0 & k = 2,3,4, \cdots \end{cases}$$

If

$$\phi(x,y) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} P_{n,k} x^n y^k \qquad (12)$$

it is readily deduced from Eqs. (10) and (11) that this generating function is given by

$$\phi(x,y) = \frac{1 + pxy}{1 - qx - px^2y} \qquad (13)$$

and the expansion of this function in a double series, Eq. (12), immediately yields the result, Eq. (6). The details are strictly algebraic and are left to the interested reader.

Although several difference equations can be found for the first problem, these are somewhat more complicated and will not be demonstrated here.

The problems can be generalized in several directions. The simplest of these is that of requiring a turnaround time of R, rather than one, days. That is, after each successful countdown, R days must elapse before the next trial may be attempted. The results can be obtained as before and in this case one obtains for Problem (1),

$$P_{n,k} = \binom{(n-1) - R(k-1)}{k-1} p^k q^{(n-1) - (R+1)(k-1)} \qquad (14)$$

which reduces to Eq. (1) for $R = 1$.

The answer for Problem (2) is somewhat more complicated to analyze directly as a little contemplation will show, although it can be done. The difference equation for this case becomes

$$P_{n,k} = qP_{n-1,k} + pP_{n-(R+1)(k-1)} \qquad (15)$$

As can be seen from this equation, the major difficulty is that it is now of order $(R + 1)$, in n rather than 2, and therefore presents additional difficulties.

E. Accuracy Analysis of Minimum-Energy Descent From an Elliptical Orbit

A. Kliore

1. Introduction

The problem of determining the minimum velocity-increment maneuver to descend from any point on an elliptical orbit with a specified entry angle was discussed in the previous volume of the *Space Programs Summary* (Ref. 12), and some results were shown for an elliptical

orbit about Mars. The two entry quantities that were determined, together with the maneuver quantities ΔV and β, are the entry velocity V_E and the true anomaly of the entry position Φ. The purpose of this study is to investigate the in-plane errors in the entry quantities V_E, γ_E, and x (position) resulting from:

(1) Imperfect knowledge of the parameters of the original elliptical orbit (a_1, e_1, θ_1) and of the gravitational constant μ of the planet.

(2) Errors in the execution of the maneuver $(\Delta V, \beta)$.

It is assumed that errors in the determination of the orbital elements affect the computation of the optimal maneuver quantities, as well as the mechanics of the descent trajectory.

Linear analysis was used throughout this study except where first-order partial derivatives vanish, in which cases second-order terms were included.

A digital computer program using the results of this study is described in Section 3. The program computes the standard deviations of the entry velocity, entry angle, and entry position for each optimal descent trajectory, given the gravitational constant of the central body, the apoapsis and periapsis altitudes of the initial orbit, the entry altitude, and the standard deviations of the six error sources.

Finally, some computed results for a typical elliptical orbit about Mars are described in Section 4.

2. Mathematical Development

Uncertainties in μ, a_1, e_1, θ_1, ΔV, and β are denoted by $\delta\mu$, δa_1, δe_1, $\delta\theta_1$, $\delta\Delta V$, and $\delta\beta$, respectively. They are all assumed to be independent, Gaussian, have zero means, and standard deviations of σ_μ, σ_{a_1}, σ_{e_1}, σ_{θ_1}, $\sigma_{\Delta V}$, and σ_β, respectively.

Using conic formulas and linear expansions, resorting to second-order terms whenever necessary, it is possible by the repeated application of the chain rule to express the errors in the entry quantities V_E, γ_E, and x (represented by δq_k, $k = 1, 2, 3$) in terms of the six original error components as follows:

$$\delta q_k = \sum_i \left(\frac{Dq_k}{D\alpha_i}\right) \delta\alpha_i + \sum_i \left(\frac{D^2 q_k}{D\alpha_i^2}\right) \delta\alpha_i^2$$
$$+ \sum_{i \neq j} \left(\frac{D^2 q_k}{D\alpha_i D\alpha_j}\right) \delta\alpha_i \delta\alpha_j \qquad (1)$$

where α_i, $i = 1, 2 \cdots 6$ represent μ, a_1, e_1, θ_1, ΔV, and β, respectively, and the coefficients $(Dq_k/D\alpha_i)$, $(D^2q_k/D\alpha_i^2)$, and $(D^2q_k/D\alpha_i D\alpha_j)$ are complicated functions of certain partial derivatives computed on the standard descent trajectory.

Since all of the $\delta\alpha_i$ are statistically independent and have Gaussian distributions, the following is true:

$$E\left[\delta q_k\right] = \sum_i \frac{D^2 q_k}{D\alpha_i^2} \sigma_{\alpha_i}^2 \qquad (2)$$

$$\sigma_{\delta q_k}^2 = E\left[\delta q_k^2\right] - E^2\left[\delta q_k\right]$$

$$= \sum_i \left(\frac{Dq_k}{D\alpha_i}\right)^2 \sigma_{\alpha_i}^2 + 2 \sum_i \left(\frac{D^2 q_k}{D_i^2}\right)^2 \sigma_{\alpha_i}^4$$

$$+ \sum_{i \neq j} \left(\frac{D^2 q_k}{D\alpha_i D\alpha_j}\right)^2 \sigma_{\alpha_i}^2 \sigma_{\alpha_j}^2 \qquad (3)$$

Because of the presence of second-order terms in Eq. (1), the distributions of the δq_k are no longer Gaussian and their means are not zero. However, it will be shown later that in the case of the numerical example, the means are of negligible magnitude and the standard deviations can be treated as representing a Gaussian distribution.

3. Computer Program

The computations described in Eqs. (2) and (3) were programmed for a digital computer. The program accepts as input the periapsis and apoapsis altitudes of the original orbit (km), the gravitational constant of the planet (km³/sec²), the radius of the planet (km), the entry radius (km), and the six standard deviations σ_μ (km³/sec²), σ_{a_1} (km), σ_{e_1}, $\sigma_{\Delta V}$ (expressed as a percentage of the total velocity increment ΔV), σ_β (deg), and σ_{θ_1}, which is expressed in terms of the uncertainty in the knowledge of the Cartesian position coordinates resulting from uncertainties in orbit determination. The quantity that is specified in the input is the rms value of the standard deviations of the x, y, and z coordinates, and σ_{θ_1} is then computed by dividing this quantity by the radial distance to the center of the planet.

The program first computes the optimal descent maneuver for points on the original orbit spaced 10 deg apart, for a range of entry angles from 5 to 85 deg, and for each resulting trajectory the program then computes the mean and standard deviation of each of the entry quantities V_E, γ_E, and x.

4. Results

The program described in the preceding section was used to compute the accuracy of descent trajectories from an orbit about Mars having an apoapsis altitude h_a of 10,000 km and a periapsis altitude h_p of 1,800 km. Based on estimates of the accuracy of the existing orbit determination techniques, the 1-σ uncertainties in μ, a_1, and e_1 were assumed to be 10 km³/sec², 1.0 km, and 4×10^{-6}, respectively. The 1-σ uncertainty in the knowledge of the Cartesian position coordinates was assumed to be 1 km in each coordinate, or an rms value of $\sqrt{3}$ km. The 1-σ execution errors were assumed to be 0.5% of ΔV for the velocity increment, and 0.8 deg for the pointing error (assumed attitude control accuracy).

Fig. 1 shows the standard deviation of the entry velocity σ_{V_E} as a function of the true anomaly of the deorbit point θ_1 for nine values of γ_E, ranging from 5 to 85 deg in steps of 10 deg. (See Table 1: constants for Figs. 1–13.)

Similarly, Figs. 2 and 3 show the standard deviations of the entry angle σ_{γ_E} and entry position σ_x, respectively, plotted in the same manner. It may be noted that the curves in Figs. 1 and 2 are plotted for θ_1 between 0 and 180 deg, and in Fig. 3 they are plotted for θ_1 between 0 and 360 deg. The reason for this is that σ_{V_E} and σ_{γ_E} are symmetrical about $\theta_1 = 180$ deg, but σ_x is not, because of the difference in the total true anomaly of the descent trajectory for $\theta_1 < 180$ deg, and $\theta_1 > 180$ deg.

In Fig. 4, σ_{V_E} is plotted as a function of γ_E for values of θ_1 of 0, 10, 30, 60, 90, 120, 150, 160, and 180 deg to show the dependence of σ_{V_E} on the entry angle γ_E. Similarly Fig. 5 shows σ_{γ_E} as a function of γ_E, and Figs. 6 and 7 show σ_x plotted vs γ_E with θ_1 as a parameter.

To investigate the dependence of the standard deviations σ_{V_E}, σ_{γ_E}, and σ_x upon the pointing error, a computation was carried out with $\sigma_\beta = 0$ (no pointing error). The standard deviations of the resulting errors in the entry quantities V_E, γ_E, and x are plotted as functions of

the true anomaly at release θ_1 in Figs. 8, 9, and 10, respectively. It may be noted that the magnitudes of all standard deviations are reduced by a significant amount.

Table 1. Constants for Figs. 1–13

Parameter	Value
h_a, km	10,000
h_p, km	1,800
σ_μ, km^3/sec^2	10
σ_{a_1}, km	1
σ_{e_1}	4×10^{-6}
σ_{θ_1}	$\sqrt{3}/\rho$
$\sigma_{\Delta V}$, km/sec	$0.005\,\Delta V$
σ_β, deg	0.8

Finally, a computation was made with both execution errors set to zero, i.e., with only the effects of the orbit determination errors present. The resulting standard deviations in the entry quantities V_E, γ_E, and x are shown in Figs. 11, 12, and 13, respectively. Upon examination of these graphs, it is evident that the magnitudes of all standard deviations are drastically reduced.

Several conclusions can be drawn on the basis of these results. It is first of all evident that the effect of the uncertainties in orbit determination on the dispersion of entry condition errors is far smaller than the effect of execution errors, and is, in fact, almost negligible in the case of the example used in this study.

In almost all cases, the errors are least when the maneuver is performed at apoapsis of the original orbit, which is also the point at which the required velocity increments are at a minimum.
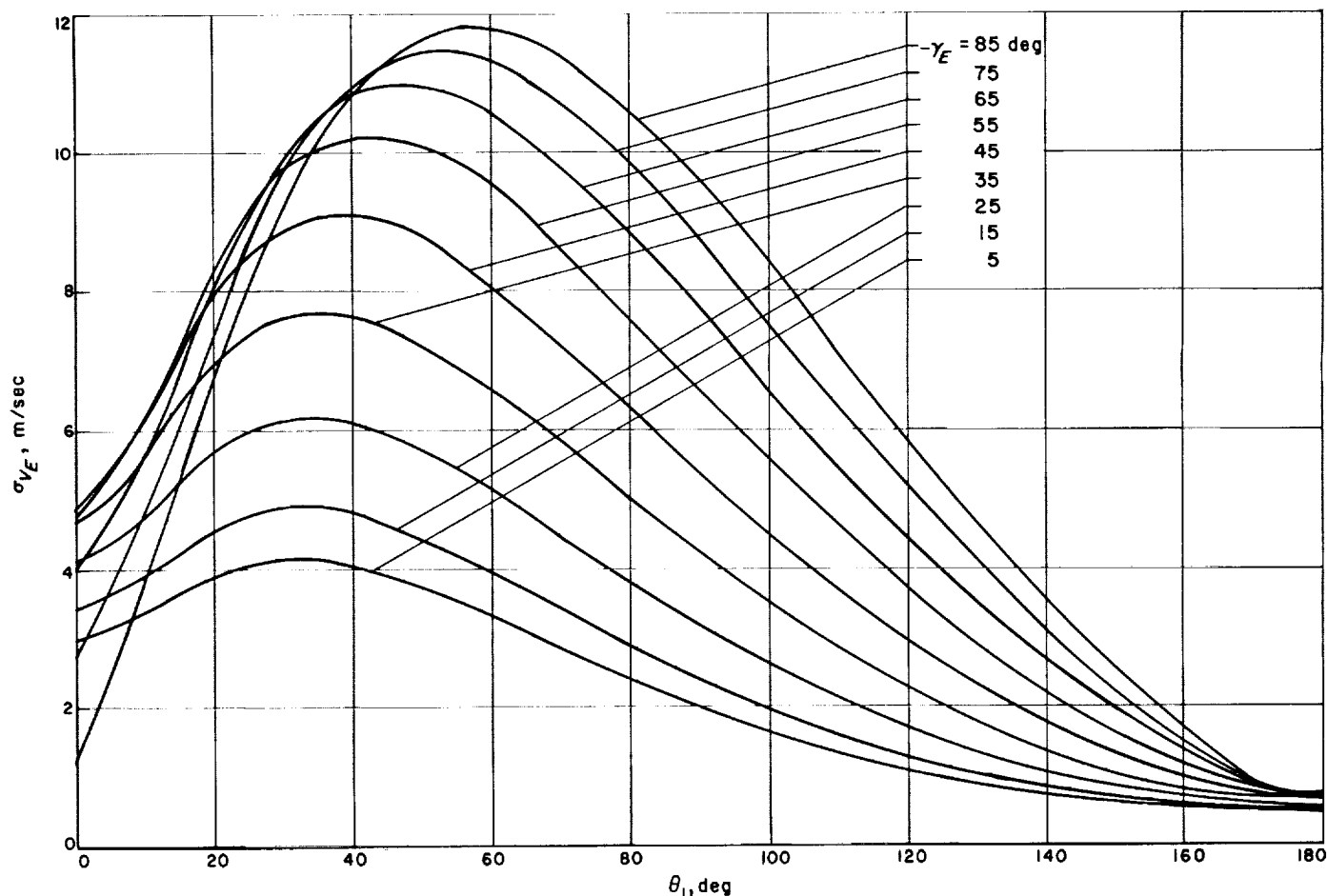
Fig. 1. Standard deviation of entry velocity vs true anomaly of deorbit point
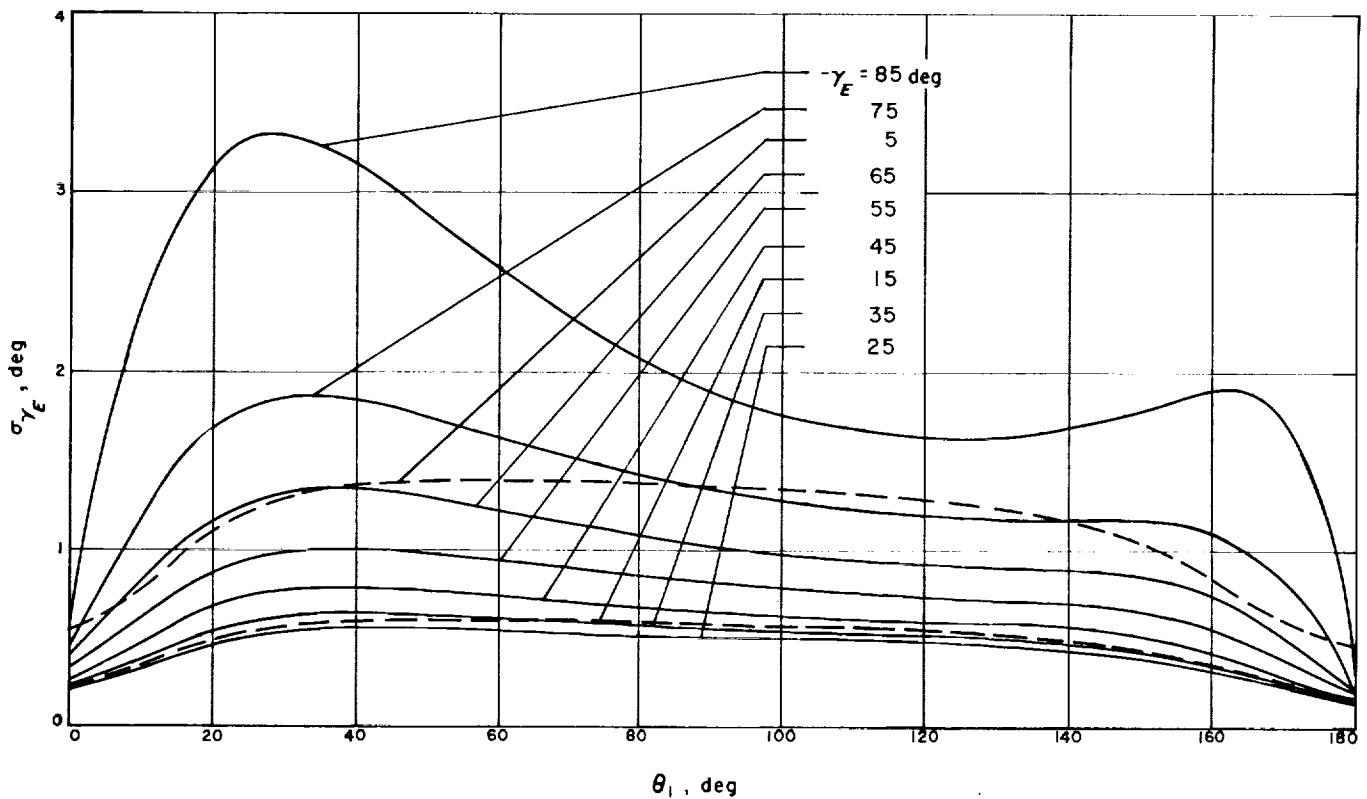
Fig. 2. Standard deviation of entry angle vs true anomaly of deorbit point
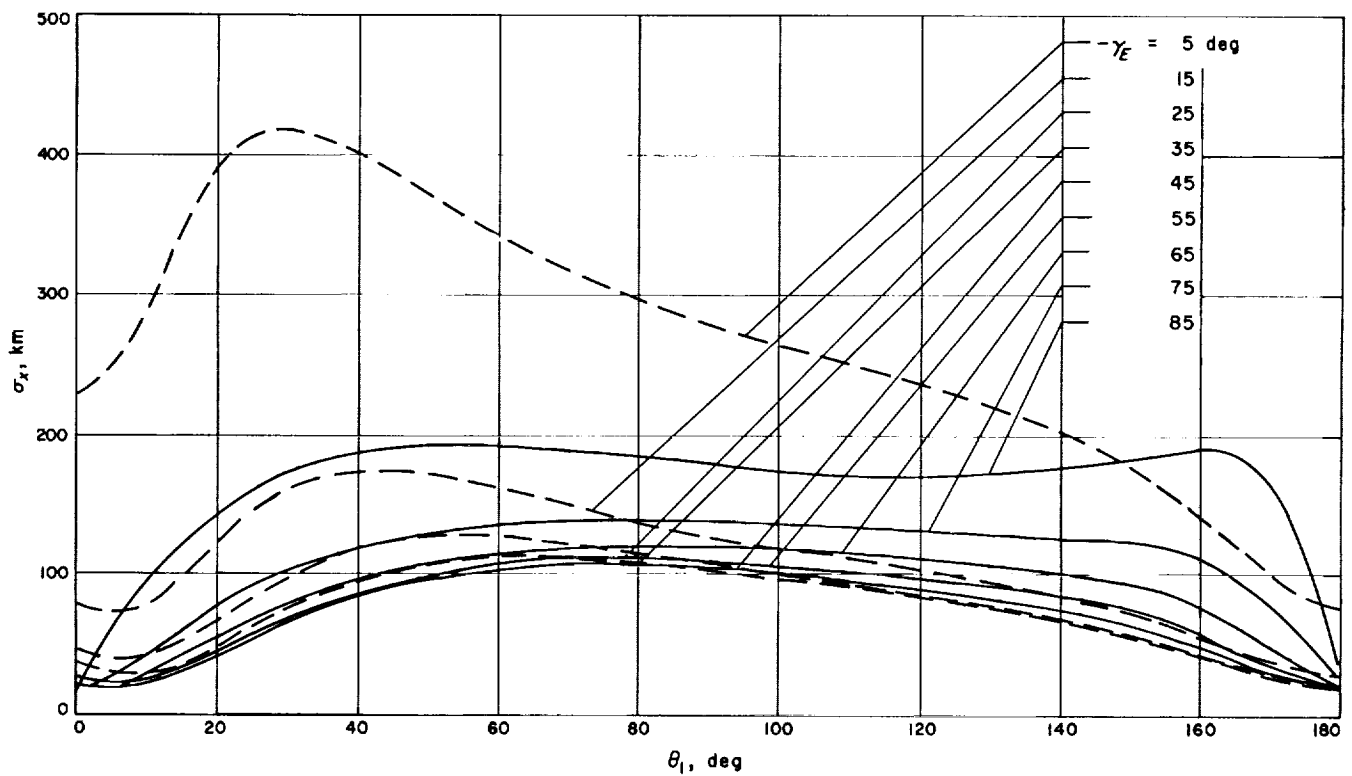
Fig. 3. Standard deviation of entry position vs true anomaly of deorbit point
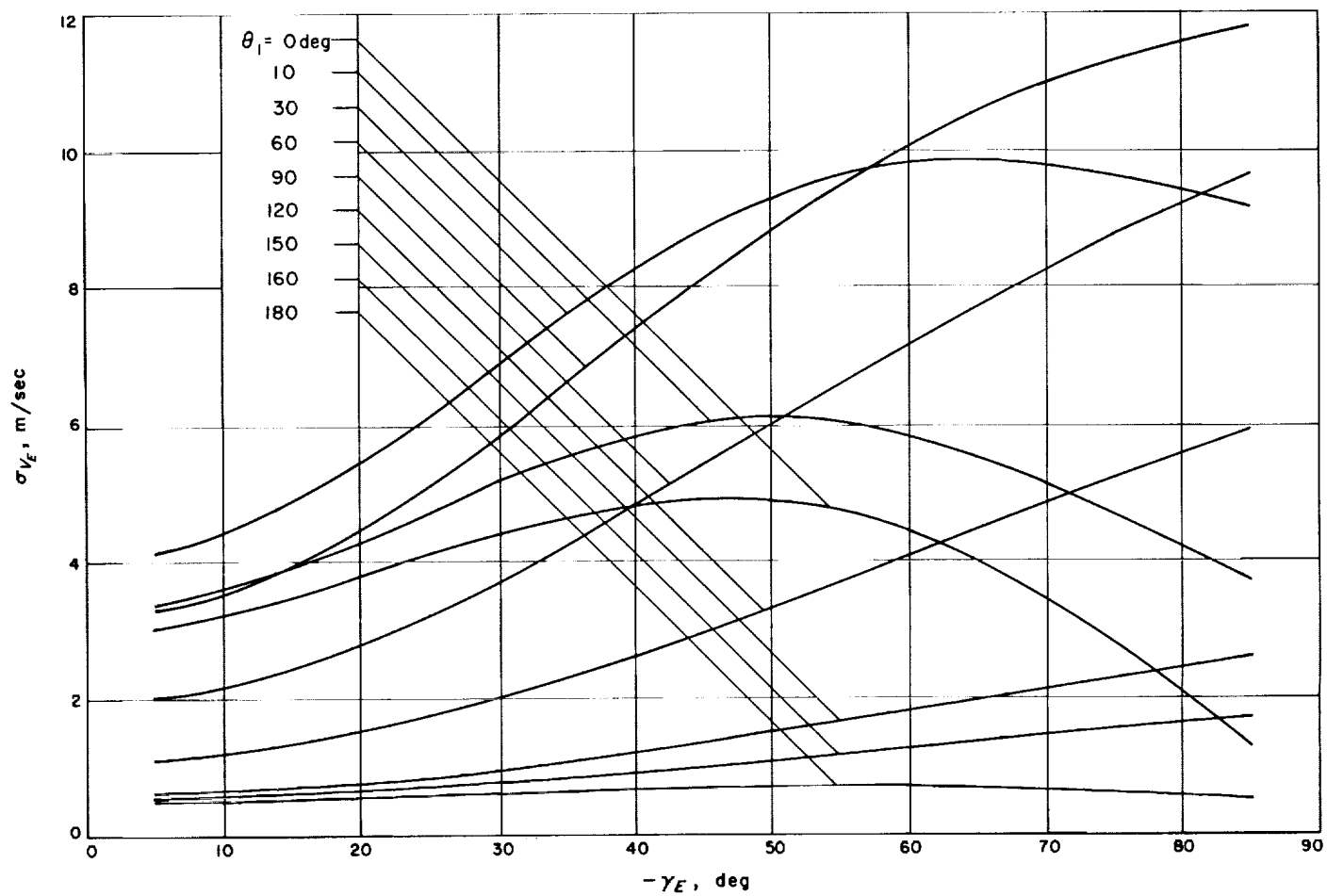
Fig. 4. Standard deviation of entry velocity vs entry angle
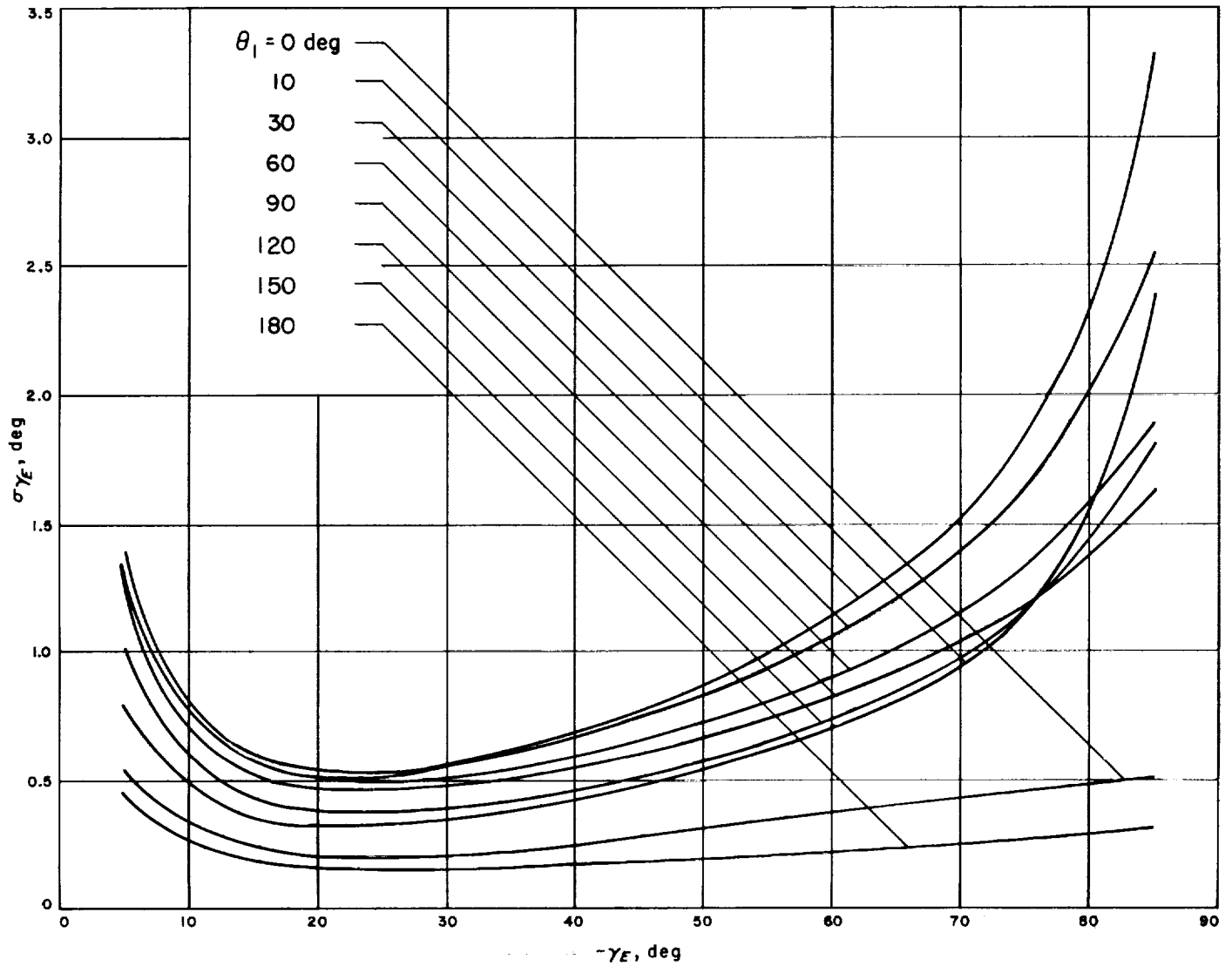
Fig. 5. Standard deviation of entry angle vs entry angle
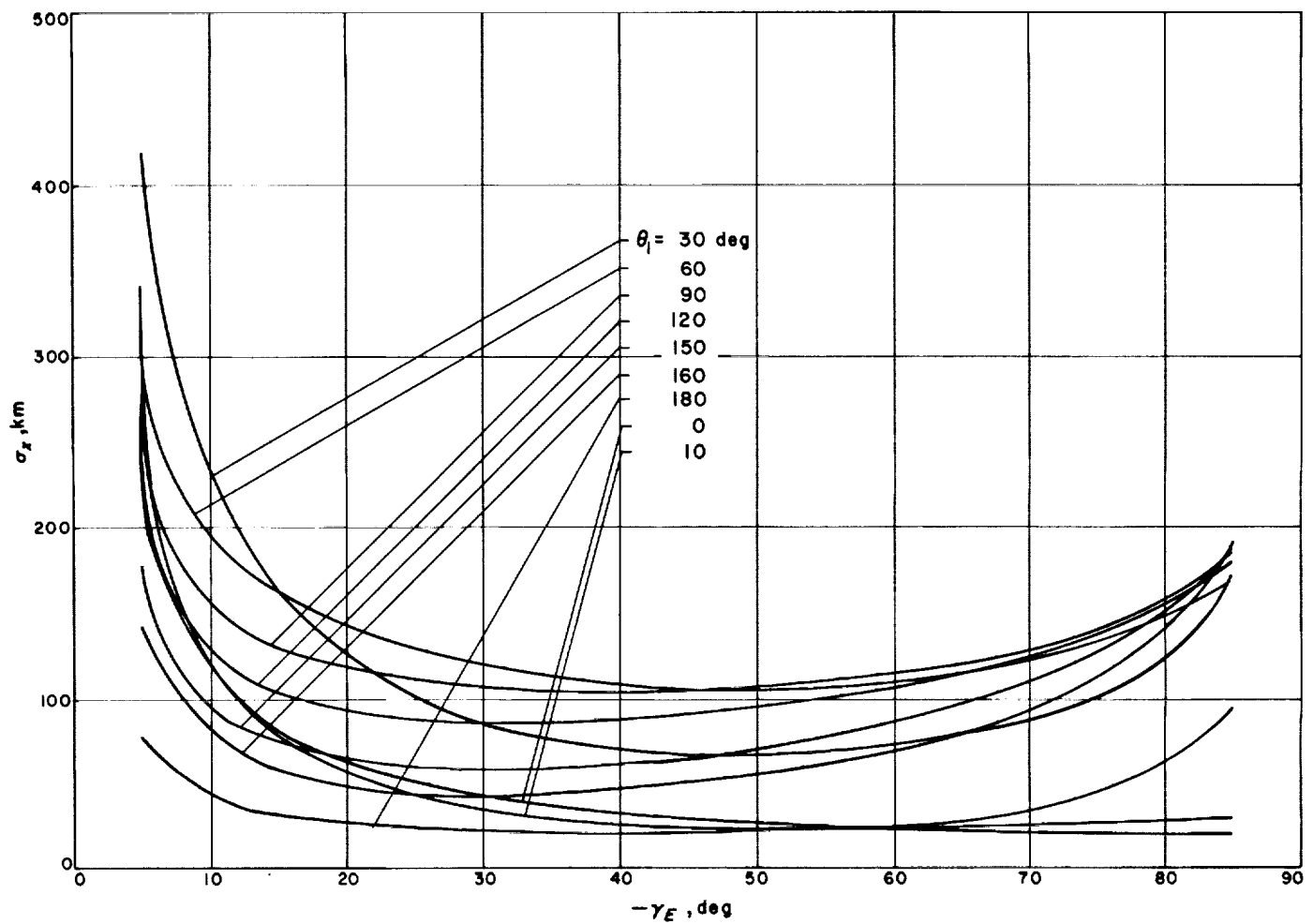
Fig. 6. Standard deviation of entry position vs entry angle, $\theta_1 < 180$ deg
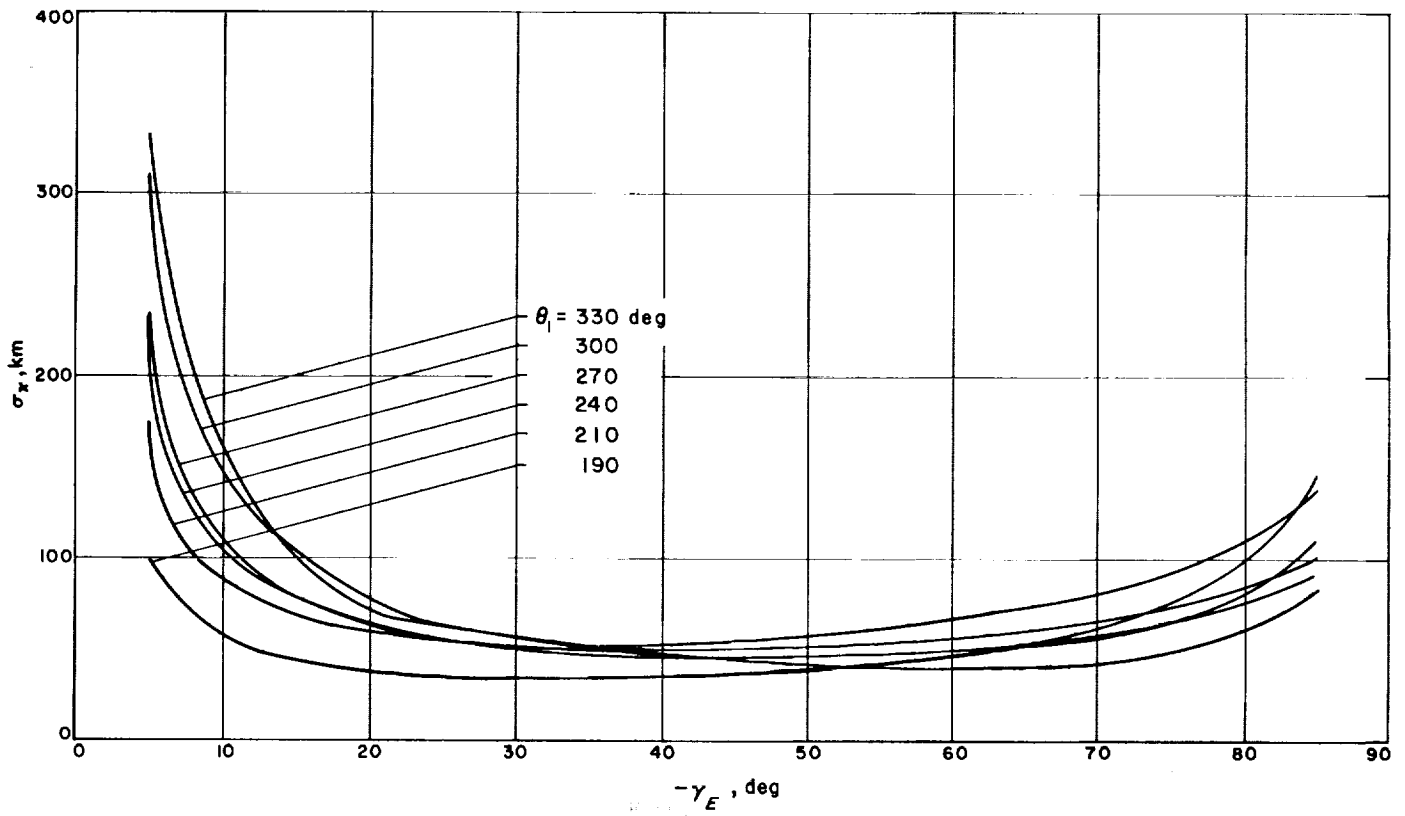
Fig. 7. Standard deviation of entry position vs entry angle, $\theta_1 > 180$ deg
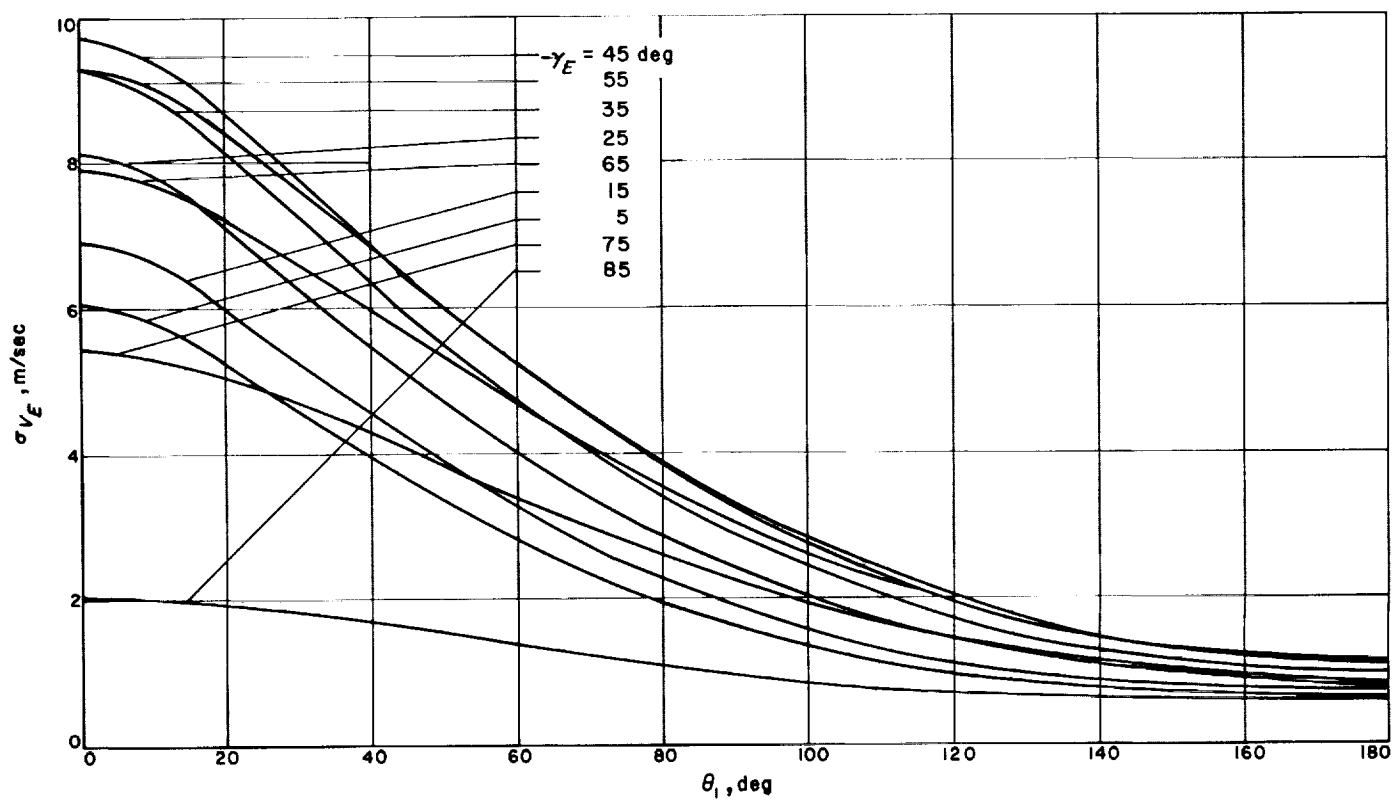
Fig. 8. Standard deviation of entry velocity vs true anomaly of deorbit point, $\sigma_\beta = 0$
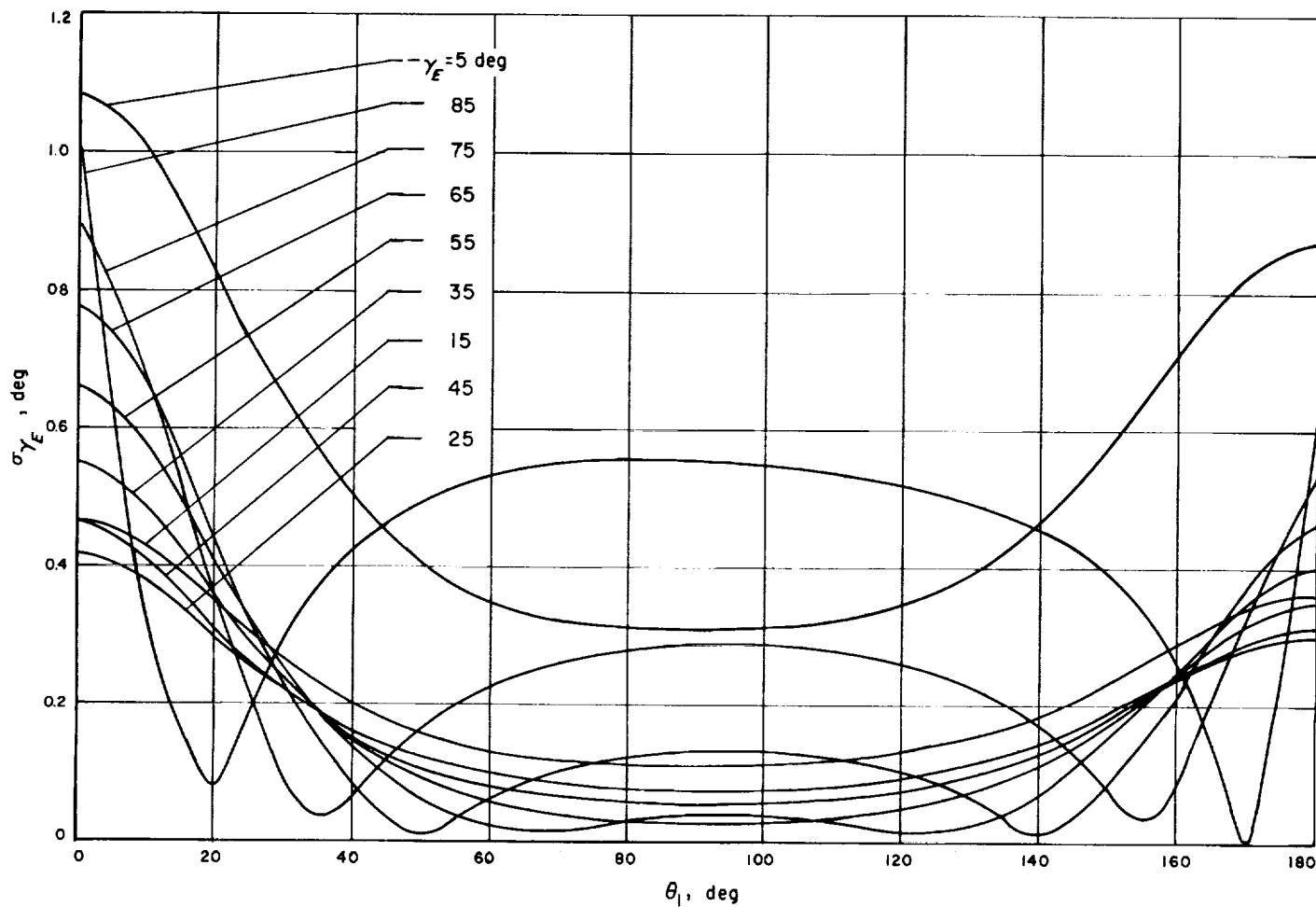
Fig. 9. Standard deviation of entry angle vs true anomaly of deorbit point, $\sigma_\beta = 0$
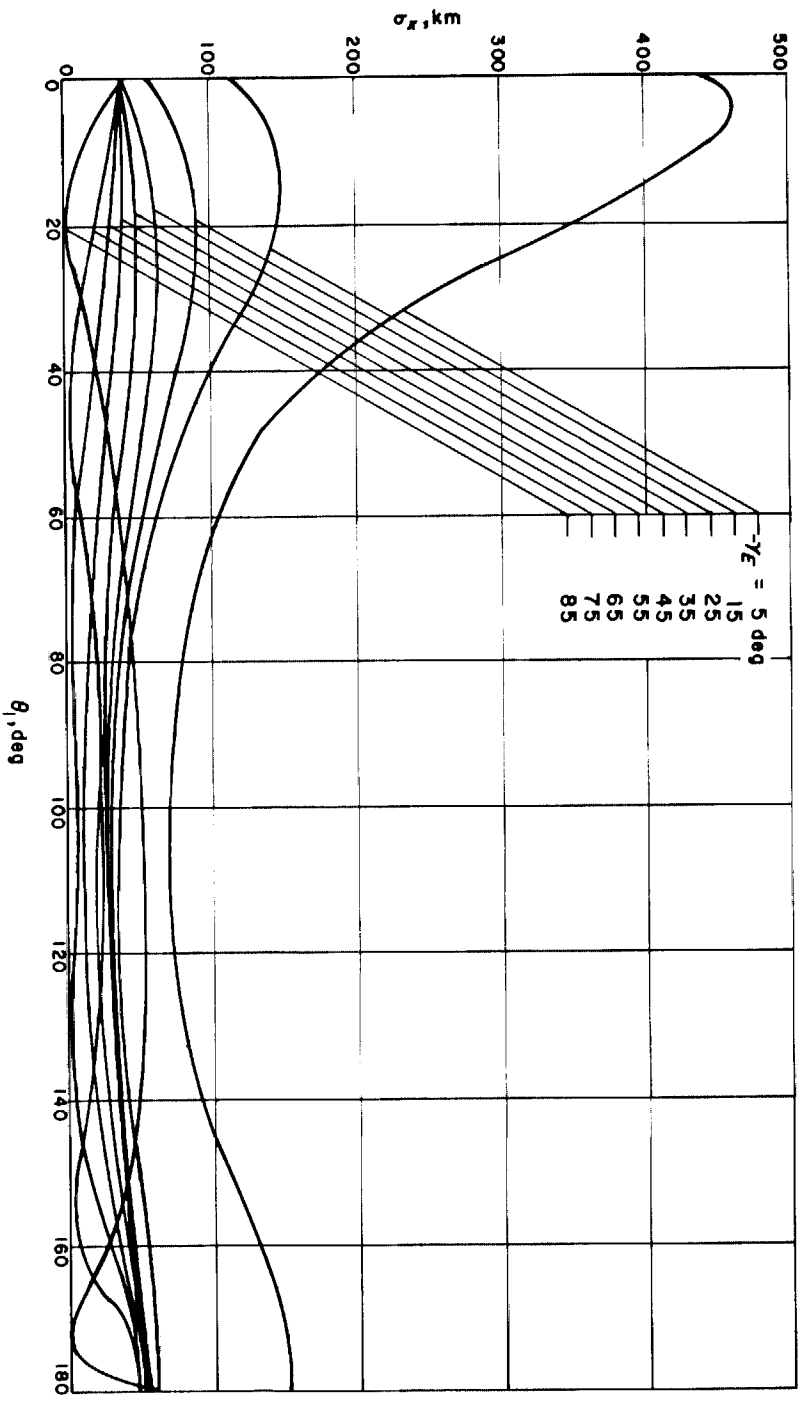
Fig. 10. Standard deviation of entry position vs true anomaly of deorbit point, $\sigma_\beta = 0$
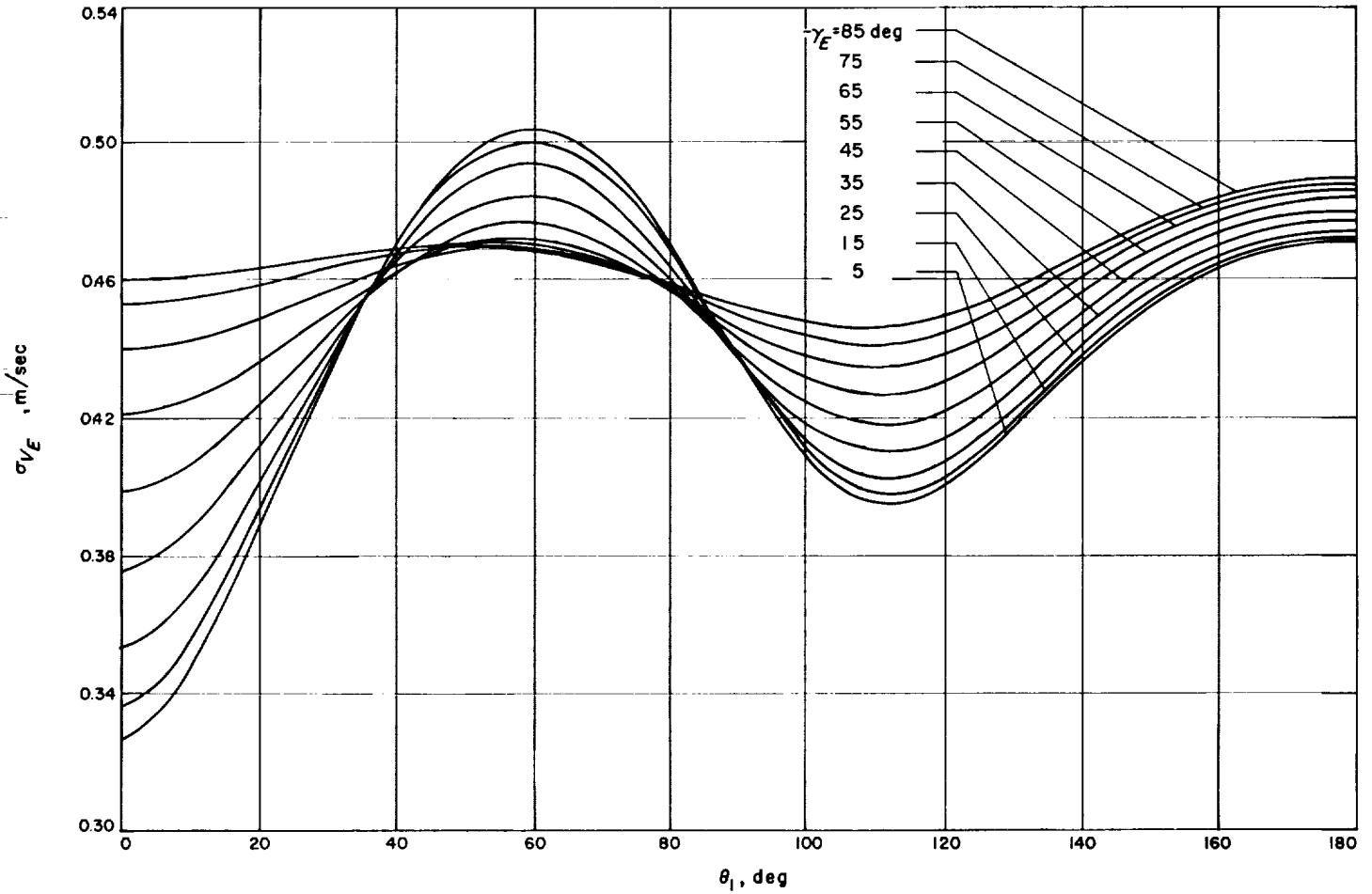
Fig. 11. Standard deviation of entry velocity vs true anomaly of deorbit point, $\sigma_\beta = 0$, $\sigma_{\Delta V} = 0$
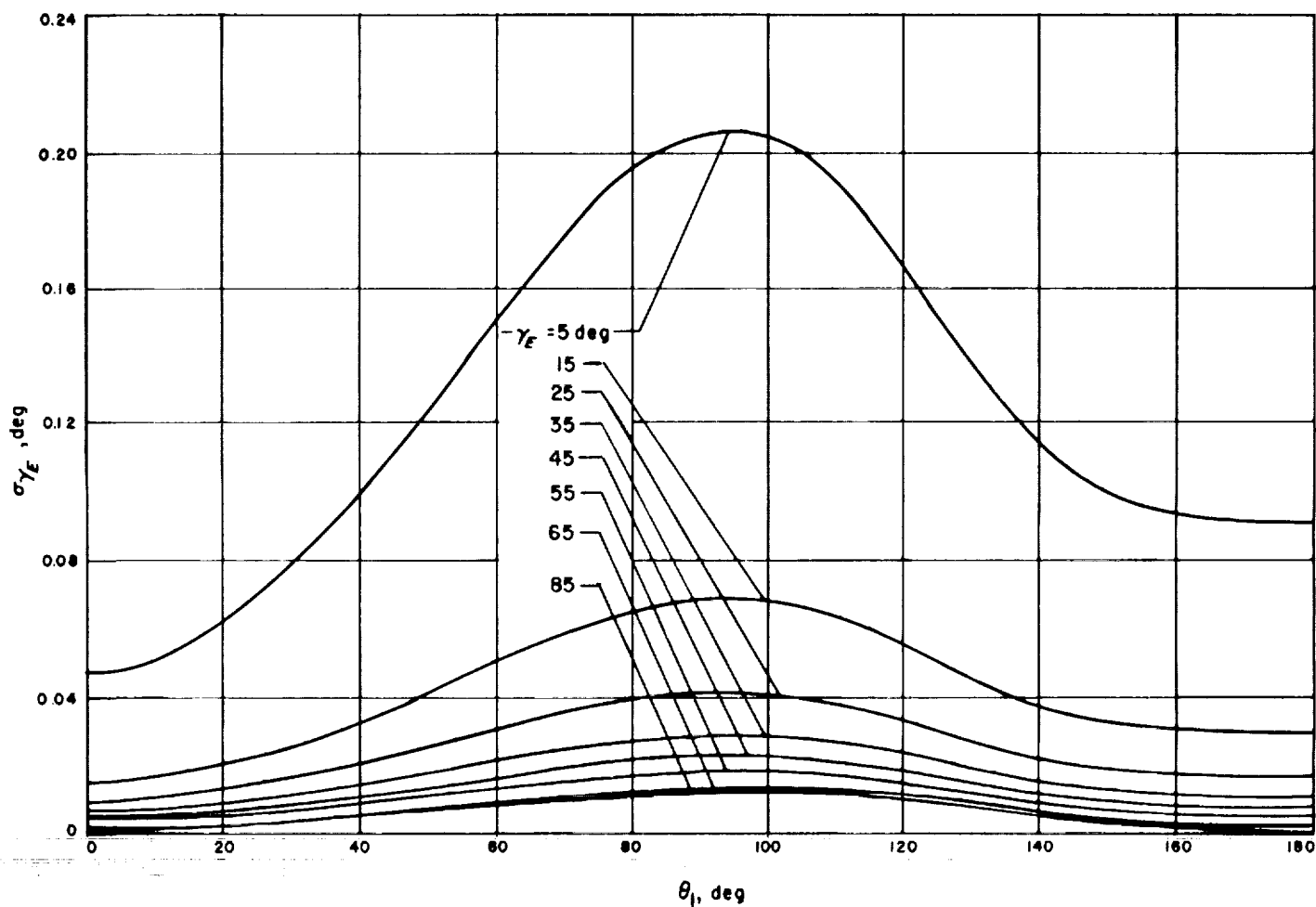
Fig. 12. Standard deviation of entry angle vs true anomaly of deorbit point, $\sigma_\beta = 0$, $\sigma_{\Delta V} = 0$
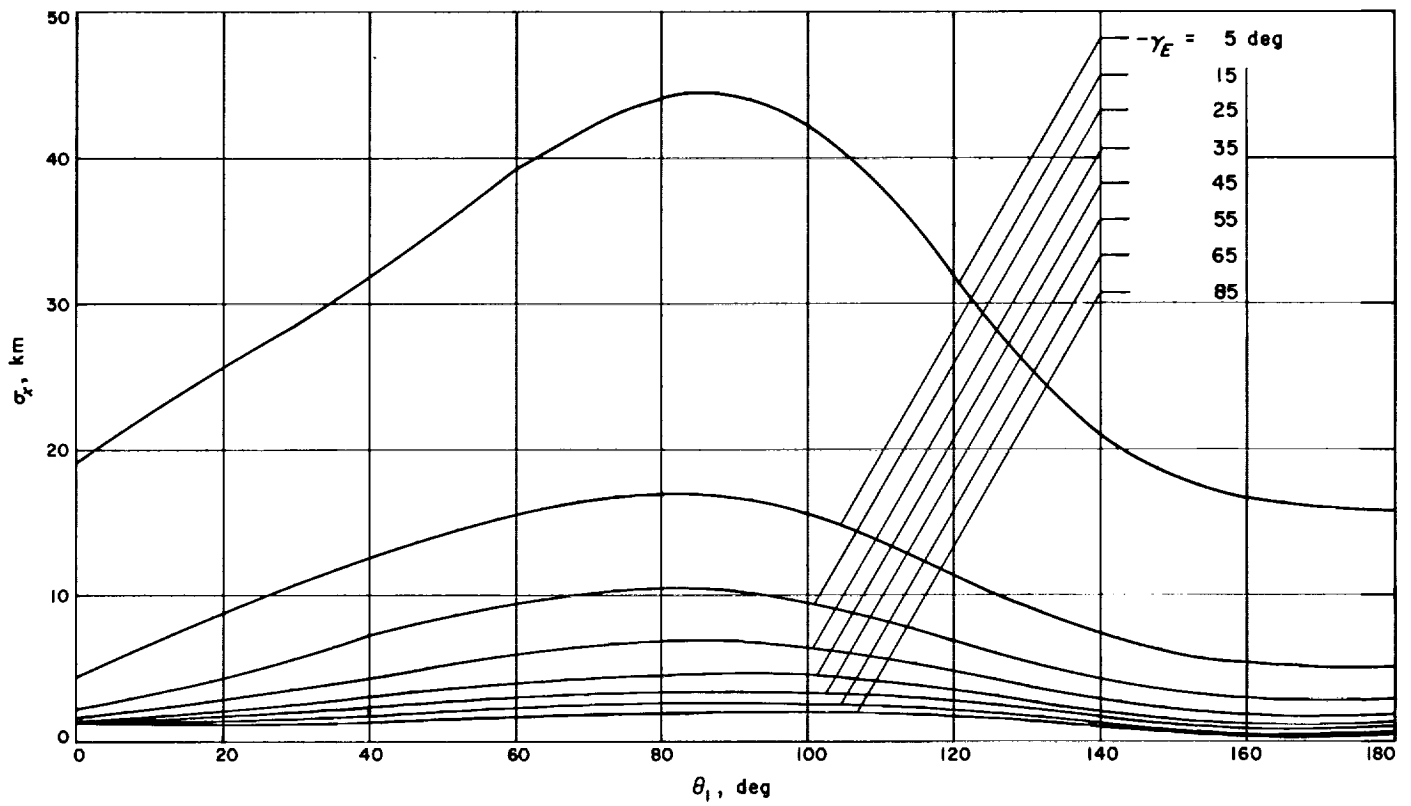
Fig. 13. Standard deviation of entry position vs true anomaly of deorbit point, $\sigma_\beta = 0$, $\sigma_{\Delta V} = 0$

References

1. Broucke, R., "Orbites Periodiques Symetriques Dans le Probleme Restreint Plan (Systeme Terre-Lune)," Doctoral Dissertation, University of Louvain, Belgium, 1963.

2. Broucke, R., *Regularizations of the Plane Restricted Three-Body Problem*, Technical Report, Jet Propulsion Laboratory, Pasadena, California, to be published.

3. Lemaitre, G., "Regularization of the Three-Body Problem," *Vistas in Astronomy*, edited by A. Beer, Vol. 1, pp. 202–215, 1955.

4. Arenstorf, R. F., "Periodic Trajectories Passing Near Both Masses of the Restricted Three-Body Problem," presented at the XIVth International Astronautical Congress, Paris, France, 1963.

5. Thiele, T. N., "Recherches Numeriques Concernant des Solutions Periodiques d'un Cas Special du Probleme des Trois Corps," *Astronmische Nachichten*, Vol. 3289, pp. 1–10, 1895.

6. Deprit, A., Broucke, R., "Regularisation du Probleme Restreint Plan des Trois Corps par Representations Conformes, *Icarus*, Pergamon Press, New York City, N. Y., Vol. 2, pp. 207–218, December, 1963.

7. Pontryagin, et al., *The Mathematical Theory of Optimal Processes*, Interscience Publishers, New York City, New York, 1962.

8. Breakwell, J. V., "The Optimization of Trajectories," *Journal of the Society for Industrial and Applied Mathematics*, Vol. 7, pp. 215–247, 1959.

9. Pfeiffer, C. G., Technical Report No. 32-566, On *the Second Variation of an Optimal Trajectory*, to be published, Jet Propulsion Laboratory, Pasadena, California.

10. Lovitt, W. V., *Linear Integral Equations*, Dover Publications, New York City, New York, 1950.

11. Bliss, G. A., *Lectures in the Calculus of Variations*, University of Chicago Press, Chicago, Illinois, 1946.

12. Kliore, A., "Descent From an Elliptical Orbit With a Minimum Velocity Increment," SPS 37-25, Vol. IV, pp. 7–9, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

GUIDANCE AND CONTROL DIVISION

II. Spacecraft Electrical Power

A. Power Sources

P. Rouklove, D. W. Ritchie, and R. K. Yasui

1. Solar Energy Thermionic Electrical Power Supply Development, P. Rouklove

a. Life testing. The completion of a successful life test of five solar energy thermionic (SET) converters was reported in Ref. 1. The results of the initial acceptance test and the life test are presented in Tables 1 and 2, respectively. The life test was performed in independent units which were carefully controlled and locked to prevent access by unauthorized personnel. All five converters performed satisfactorily for over 3000 hr of near-continuous operation at an emitter temperature of approximately 2000°K. The input power to the converters was kept constant throughout the test period, allowing emitter temperature and output power to vary during the test. Such conditions were considered close to those encountered during the natural mode of operation of the SET converter. These fluctuations were presented in Ref. 1, the most significant being the apparent decrease in the output power observed in Converter VI-TEP-1, which dropped from 7.3 to 4.4 w after 2100 hr of operation. Since most of the observed decreases in output power could be correlated with an accompanying drop in emitter temperature, it was suggested that these changes could be attributed to changes in efficiency of electron bombardment or to a change in heat transfer through the converter. Throughout the testing period, the vacuum in the test chamber was maintained below 2×10^{-7} mm Hg. Ion-type pumps were employed in this operation to avoid reactions due to possible oil backstreaming.

Table 1. Converter acceptance test results

Converter	Output voltage, v	Output current, amp	Output power density, w/cm²	Input power, w	Temperature, °C				
					Cesium	Radiator	Collector	Seal	Emitter (observed)
VI-S-14	1.0	26.0	26.0	407	407	593	—	604	1710
VI-S-18	1.0	13.5	6.75	328	374	572	—	654	1655
VI-TEP-1	1.0	8.5	4.25	310	367	564	—	661	1655
VII-S-4	0.8	19.1	7.64	335	368	551	590	697	1655
VII-S-5	0.8	18.0	7.2	365	345	540	583	692	1655

Table 2. Converter life test results

Converter	Elapsed time, hr	Output power density, w/cm²	Input power, w	Temperature, °C		
				Cesium	Seal	Emitter (observed)
VI-S-14	0	7.0	344	355	614	1655
	1046.6	5.7	378	360	644	1655
	2010	5.3	378	364	652	1655
	3013	5.25	378	364	649	1655
VI-S-18	0	6.65	370	360	670	1655
	1049	6.2	376	316	653	1670
	2047	5.8	376	354	633	1660
	2992	6.0	376	350	654	1660
VI-TEP-1	0	4.4	300	348	634	1655
	1016.8	3.65	284	348	637	1655
	2021	2.65	300	348	637	1635
	3010.4	2.5	300	348	637	1635
VII-S-4	0	7.8	430	357	642	1655
	1043	7.6	430	362	661	1650
	2051.2	7.68	430	358	649	1650
	3026	8.08	430	363	650	1645
VII-S-5	0	7.2	350	345	659	1660
	997.6	7.0	351	374	642	1650
	1988.7	6.68	351	372	642	1640
	2998.4	6.44	351	383	633	1635

After the life test was completed, the converters were parametrically tested at JPL. The results of these tests are presented in Fig. 1. The "acceptance test point" indicates the original condition of the converter before the start of the life test (at an observed emitter temperature of 1655°C). From the JPL test results, it appears that the correction initially applied by the manufacturer to the recording of the emitter temperature (50°C) was excessive. Such a correction was felt necessary to compensate for the difference in temperature between a blackbody cavity and a hohlraum with a 4:1 depth-to-diameter ratio. When tested at JPL, Converter VI-TEP-1, which exhibited a 40% change in output power during the life test, indicated an output power of 10 w at 1 v and a 1700°C emitter temperature and 27.5 w at 0.4 v and a 1700°C emitter temperature. During the life test, the

same converter produced an output power of 7.5 w at the initiation of the test and 4.4 w at the end. These data, presented in Fig. 1(c), point to a decrease in the emitter temperature, which at the end of the test had probably dropped below 1900°K. Fig. 2 presents a typical current–voltage curve of a SET converter in which a difference in emitter temperature of 30°C resulted in a difference of 15 w in output power. This difference emphasizes the care required for accurate measurement of the emitter temperatures in SET converters.

The occasional interruptions observed during the test were attributed primarily to auxiliary component failures such as in vacuum tubes, transformers, relays, etc., or to a drop in water pressure. This points out the necessity for oversize components for reliable and trouble-free operation. An examination of the electron-beam gun used as the emitter heater seems to indicate that the observed output power degradation is probably the result of: (1) a misalignment in the gun structure due to excessive thermal deformations, (2) a power leakage loss along the insulators due to metallic coatings produced by evaporation, or (3) a change in surface emissivity of the parts facing the converter (also produced by metal evaporation). Such a coating could change the thermal balance of the converter during operation. It is clear that, for any life test of long duration, special care must be taken in the design of an adequate electron-beam gun. Furthermore, since the emitter temperature is such an important parameter in testing, operation at a constant input power requires that the measured power be truly representative of the actual power applied to the converter (not including extraneous and variable heat losses in the support structure, insulators, etc.). A desirable feature for any life-test device would be the inclusion of terminals to implement more versatile measurements, such as complete current–voltage characteristics, the collector work function, dynamic ac testing, etc., which will allow a

Fig. 1. Results of JPL parametric tests of four SET converters

Fig. 2. Typical current–voltage curve of a SET converter

were available prior to the initiation of the life test, both at the manufacturer and at JPL. The other converters of Series VI and the converters of Series VII were manufactured in 1963, having the benefit of improved technology. These converters showed no tendencies for power degradation; in fact, Converter VII-S-4 showed an increase in output power at the end of the life test.

b. Converter shielding. A better estimate of the power lost by radiation from the hot portion of the emitter supporting sleeve was obtained. This part was shielded with a tantalum multilayer radiation shield during laboratory testing of the converters using the electron-beam gun for heating. The test results (Table 3) indicated that a substantial gain in efficiency was obtainable. With this shielding the operative conditions of the converter in a generator are more closely approached. A 1% net gain in efficiency of the converter is observed, corresponding to a decrease of 8°C in the emitter temperature required to obtain similar output and also to a decrease of 20 w in the input power requirement. Furthermore, the protection of the sleeve from contamination by extraneous ambient impurities is increased. Improvements in the shielding methods are now being investigated at JPL.

c. Evaluation of the 9.5-ft D mirror. The replica mirror obtained by electroforming nickel on a 9.5-ft D nickel master was tested at the Table Mountain, California, solar test site. The mirror was mounted on a solar tracker for performance evaluation (Fig. 9, p. 35, of Ref. 2) and was tested using the modified Hartmann screen test and cold calorimeter methods. As reported in Ref. 2, the Hartmann test revealed a double focal zone. One plane originated from the central portion of the mirror, within a 7.5-ft D with a 69.17-in. focal length; the other focal plane, with a focal length of 68.25 in., was produced by an outer annulus of approximately 1-ft width.

better understanding of the observed changes with time in the converter performance. Such facilities are presently under development at JPL.

Converter VI-S-14 was the only converter in which some power degradation (approximately 5%) could be detected. This converter was especially selected for life testing since: (1) it was considered representative of the state-of-the-art in 1962; and (2) extensive test data

The Hartmann test was performed using an aluminum screen perforated by 200 holes (Fig. 3), which were

Table 3. Unshielded vs shielded emitter sleeve operation

Condition	Output voltage, v	Output current, amp	Output power, w	Input power, w	Efficiency, %	Temperature, °C				
						Cesium	Radiator	Collector	Seal	Emitter (observed)
Unshielded	1.0	11.5	11.5	262	4.39	384	510	553	571	1677
Shielded	1.0	13.5	13.5	248	5.36	384	519	563	594	1678
Shielded	1.0	11.5	11.5	241	4.77	382	514	558	589	1669

Fig. 3. 9.5-ft D mirror and aluminum screen

drilled in such a pattern that each light spot which originated from a hole was representative of 1/200th of the mirror area (51.03 in.²). The position of the reflected elliptical light pattern was observed on a polar coordinate target located at the focal plane. The slope of the mirror at the observed position was determined by trigonometry. The data were then reduced to provide: (1) the geometrical characteristics of the mirror expressed as

tangential errors, and (2) the "Hartmann factor," i.e., the fraction of energy from the mirror which is received within a specified aperture. The results of the tests are given in Table 4. As can be seen, an abrupt change in contour is indicated in the region H corresponding to 91.8 in. D. The test results are presented graphically in Fig. 4 in conjunction with data supplied by the manufacturer. From the graph, it appears that the damage occurred to the mirror prior to its shipment to JPL; a closer investigation of the data seems to indicate that the region of major tangential errors (153 deg) corresponds to one of the fixation points of the mirror. The reason for this effect is still under investigation.

Table 5 presents the cold calorimetric data obtained at four aperture diameters: 0.625, 0.825, 1.00, and 1.40 in. The data were quite reproducible and correspond to an average of more than 370 data points. The test was performed by measuring the solar energy received in the water-cooled calorimeter located at the focal zone of the mirror (Fig. 5). First, the power received directly from the mirror (open calorimeter) was recorded; then, the quartz plate to be used during the generator test was interposed and readings were made. The introduction of the plate resulted in 296-w less power being produced, observed with an aperture of 1 in., or approximately 9% as compared with a minimum of 14 to 20% observed through a Pyrex dome in previous experiments. The results of the calorimeter test are presented graphically in Fig. 6.

Misorientation data were obtained by misorienting the calorimeter in increments of 6 min of arc up, down, right, and left. The data are presented in Table 6 and

Table 4. Tangential slope errors of the 9.5-ft D mirror

Zone	Radial position, in.	Slope error of mirror in radial direction, min									
		9 deg	45 deg	81 deg	117 deg	153 deg	189 deg	225 deg	261 deg	297 deg	333 deg
A	12.745	−4.61	−3.40	−1.20	−0.73	−0.49	−3.15	−2.43	−1.21	−3.40	−4.61
B	22.074	0.00	2.30	0.23	4.61	3.68	2.76	0.46	1.15	2.99	0.23
C	28.499	3.72	−1.53	−1.53	0.44	7.23	3.06	−1.97	−0.22	−0.00	−0.22
D	33.721	5.00	−2.92	−2.50	1.67	5.83	5.21	0.00	−2.70	−3.96	−2.50
E	38.236	6.54	−5.15	−0.40	4.95	10.10	5.15	1.98	−3.96	−1.19	−3.56
F	42.271	4.81	−6.97	0.94	2.83	9.05	3.77	5.65	−7.73	−3.39	−5.84
G	45.954	6.63	−1.26	−3.04	1.79	8.07	4.84	4.84	−10.76	−3.40	−2.87
H	49.362	16.72	8.19	12.46	16.21	29.35	14.00	15.70	−7.00	0.00	1.02
I	52.550	17.86	4.38	7.96	18.19	36.05	19.16	8.12	5.68	−7.31	−2.43
J	55.556	—ᵃ	—ᵃ	−9.43	2.47	—ᵃ	12.67	12.83	24.11	−11.13	−17.15

ᵃToo diffused to record.
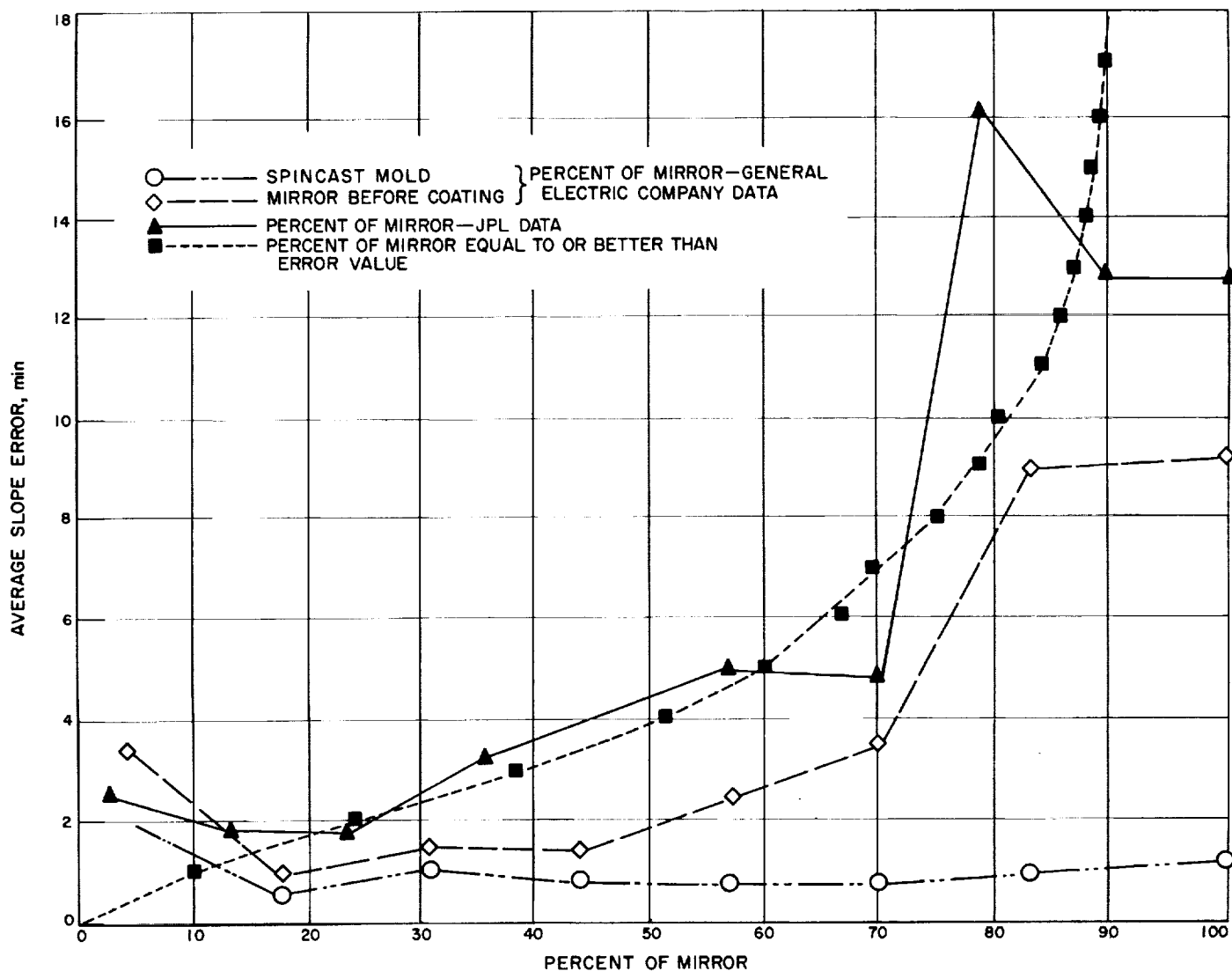
Fig. 4. Slope errors of 9.5-ft D mirror

Table 5. Results of 9.5-ft D mirror solar calorimeter tests[a]

Diameter of aperture, in.	Mirror output power, w	Efficiency of open calorimeter, %	Power available to aperture after introduction of quartz plate, w	Efficiency after introduction of quartz plate, %	Power radiating from blackbody cavity, w	Power available to generator, w	Peak power of open calorimeter (observed), w	Peak power of calorimeter after introduction of quartz plate (observed)[b], w	Concentration ratio (mirror area/aperture area)
0.625	2201.3	34.54	1960.0	30.72	179.5	1780.5	2154.2	2157.4	33,153
0.825	2787.0	43.70	2487.0	38.98	312.8	2174.2	2836.9	2747.3	15,471
1.000	3298.3	51.70	3002.3	47.06	459.6	2542.7	3374.7	3394.0	12,996
1.400	3942.8	61.80	3581.6	56.14	900.7	2680.9	4352.6	3963.6	6,631

[a]Solar intensity ranged from 87 to 101 w/ft². All data were standardized for a 90-w/ft² solar intensity.
[b]The apparent difference between the peak power data obtained before and after the interposition of the quartz plate corresponds to readings made during observations under different solar intensity inputs.
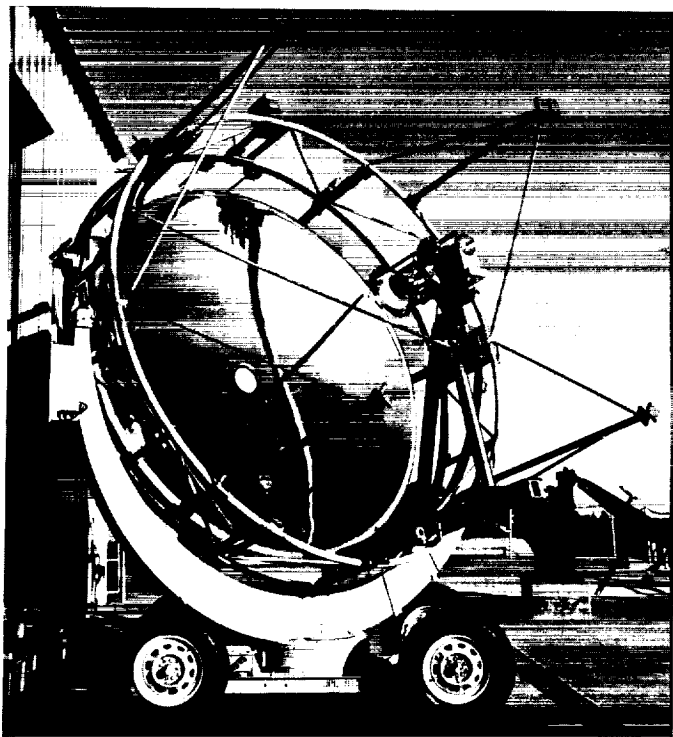
Fig. 5. 9.5-ft D mirror during solar calorimeter tests

(1) INPUT POWER TO MIRROR = 70.846 ft^2 × 90 w/ft^2
(2) POWER TO MIRROR = (1) × SHADOW FACTOR (0.9492)
(3) OUTPUT POWER FROM MIRROR (—O—) =
    (2) × REFLECTIVITY
(4) POWER AVAILABLE TO QUARTZ PLATE =
    (3) × GLASS LOSS FACTOR
(5) POWER AVAILABLE TO APERTURE AFTER
    INTRODUCTION OF QUARTZ PLATE (—□—) =
    (4) × HARTMANN FACTOR
(6) POWER AVAILABLE TO GENERATOR (—△—) =
    (5) × CAVITY RERADIATION FACTOR
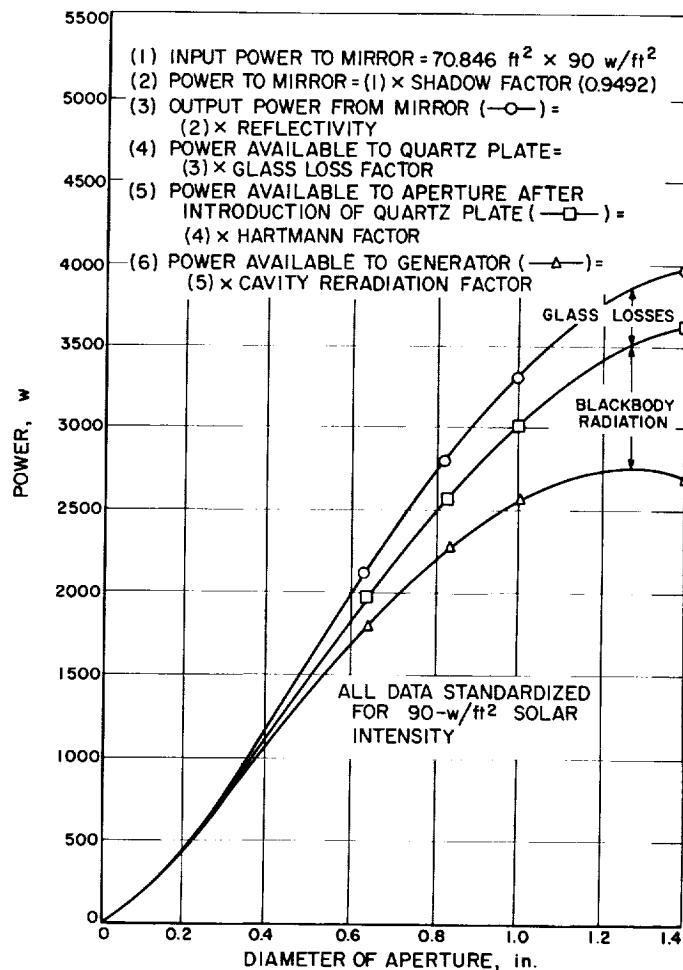
Fig. 6. Solar calorimeter test results

Table 6. Results of 9.5-ft D mirror misorientation tests

Misorientation, min	Diameter of aperture, in.	Power, w	Efficiency, %		Misorientation, min	Diameter of aperture, in.	Power, w	Efficiency %
12 up	0.625	1464.1	22.95		12 right	0.625	1623.2	25.44
	0.825	1915.7	30.03			0.825	1812.1	28.40
	1.000	2591.3	40.62			1.000	2542.3	39.85
	1.400	3430.2	53.77			1.400	3265.3	51.18
					6 right	0.625	1890.8	29.64
						0.825	2260.6	34.65
6 up	0.625	1768.7	27.72			1.000	2875.5	45.07
	0.825	2270.4	35.59			1.400	3540.2	55.49
	1.000	2924.3	45.84		12 left	0.625	1560.4	24.46
	1.400	3566.1	55.90			0.825	1953.3	30.62
						1.000	2522.1	39.53
12 down	0.625	1735.5	27.20			1.400	3251.3	50.96
	0.825	2015.5	31.59		6 left	0.625	1856.8	29.11
	1.000	2457.0	38.51			0.825	2434.7	38.16
	1.400	3117.9	48.87			1.000	2861.8	44.86
						1.400	3480.0	54.56
6 down	0.625	1961.7	30.75		0	0.625	1960.0	30.72
	0.825	2351.4	36.86			0.825	2487.0	38.98
	1.000	2822.6	44.24			1.000	3002.3	47.06
	1.400	3539.5	53.91			1.400	3581.6	56.14
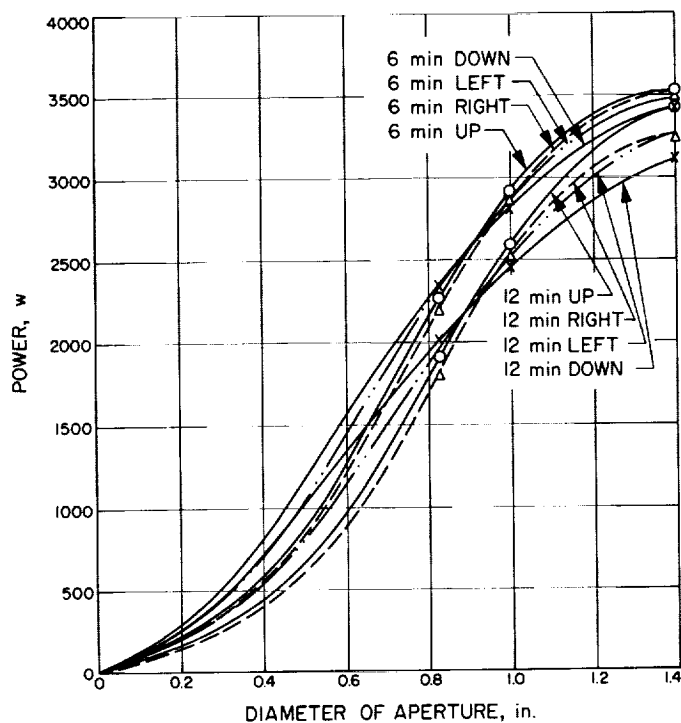
Fig. 7. Misorientation test results
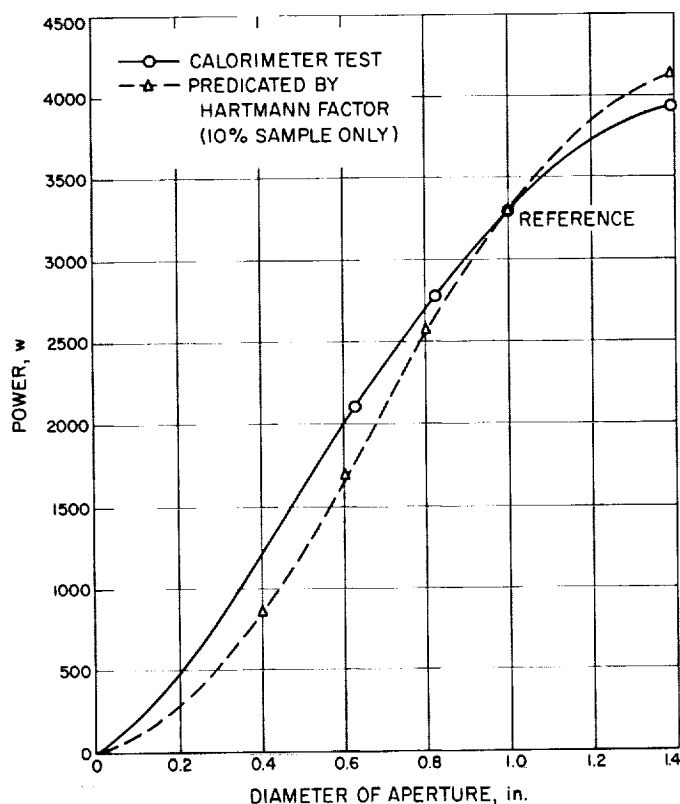
Fig. 8. Comparison of data obtained by calorimeter tests and those predicated by Hartmann factor

Fig. 7. The complete reduction of the data obtained by the Hartmann test is still in process and will give the Hartmann factor for the range of apertures corresponding to potential generator configurations. Fig. 8 illustrates the correspondence between the data obtained by direct observation with the cold calorimeter and the data predicated by the Hartmann factor as computed at a 1-in. calorimeter aperture for two angular positions (10% sample); the close correlation of the predications can be seen.

Although certain improvements are still necessary, it was decided that the mirror replica is satisfactory for use in testing of SET generators. It was observed that the 5-ft D mirror, primarily because of obscuration, could never deliver a maximum of 1200 w. Laboratory test results indicated that approximately 300 w were required to heat the emitter of a SET converter to full operating temperature. The power delivered by the 9.5-ft D mirror will heat the generator to full operation temperature. The results of further tests will be reported in subsequent issues of the SPS.

2. Solar Cell Development, D. W. Ritchie and R. K. Yasui

Programs have been initiated to study: (1) the optimization of operating characteristics of silicon solar cells for planetary exploration missions, and (2) the effect of heat sterilization on solar cell performance.

a. Solar cell optimization. It has been found that solar cells at low-level energy inputs exhibit an effective resistance of the junction, thus making the solar cell inefficient at extremely low light levels. To overcome this deficiency, the junction width and the effective resistance of the cell must be tailored for operation at a specific intensity. The solar cell series resistance at high energy levels causes excessive voltage losses and contributes to solar cell array inefficiencies. The solar cell construction can be optimized to operate at these conditions.

The effect of temperature on the current–voltage characteristics also controls the efficiency of solar cell arrays. Experimental tests have established that the open-circuit voltage decreases linearly with temperature, while the short-circuit current increases linearly. The effect of temperature on solar cell efficiency depends on the type of load (constant or maximum available).

The combined effect of temperature and intensity controls the efficiency of solar cell arrays. A study of the intensity and temperature characteristics of solar cells

can lead to the optimization of the solar cell for specific missions. The optimized solar cell could, in effect, increase the power and decrease the weight of a solar panel array for either a Mars or Venus mission.

Four companies have been developing a solderless, thin (0.008 in.), single-crystal, silicon, phosphorus-diffused, photovoltaic cell which would exhibit high efficiency and optimum operating characteristics for a specific mission. To date 300 units have been received. The solar cells are extremely fragile; therefore, special handling and testing procedures will have to be developed. The testing program will involve the measurement of the cell characteristics versus intensity and temperature, and then a comparison of these data with those of solar cells of normal configuration. The possibility of eliminating the silicon monoxide coating, which increases the absorbtivity of the cell, will also be investigated. Special soldering (or welding) methods will be studied for future solar panel array design. A prototype solar panel array will be built using the lightweight cell. Preliminary testing of the solar cells which were received shows approximately equal power output and a weight reduction of 60% compared to a 2- × 2-cm solder-dipped cell.

b. Heat sterilization of solar cells. Tests are being conducted to evaluate the effects of heat sterilization and temperature storage on solar cell assemblies. Systematic studies are being conducted on the mechanical and electrical characteristics of N/P and P/N silicon solar cells after exposure to elevated temperature environments. To date, approximately 500 solar cells of both types from two manufacturers were tested. Cells were electrically tested under tungsten and were placed on a special handling fixture instrumented with thermocouples (Fig. 9).

Trays of the cells were then placed in constant-temperature ovens for predetermined periods of time. After completing their temperature soak, the cells were removed from the oven, allowed to return to room temperature, and tested for electrical performance, again under tungsten illumination. The relative power output of the cells after exposure to elevated temperatures is shown in Fig. 10.
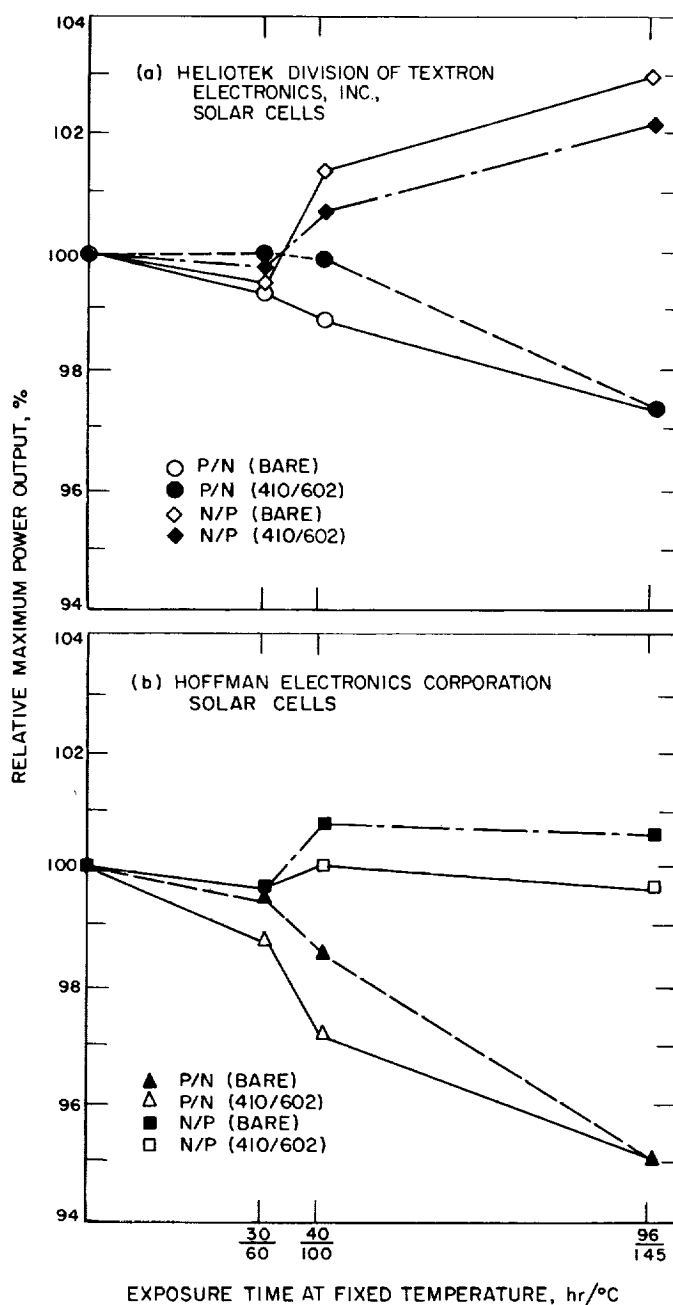
Fig. 10. Relative power output of silicon solar cells after exposure to elevated temperatures

Fig. 9. Heat sterilization test specimens mounted in special test fixture
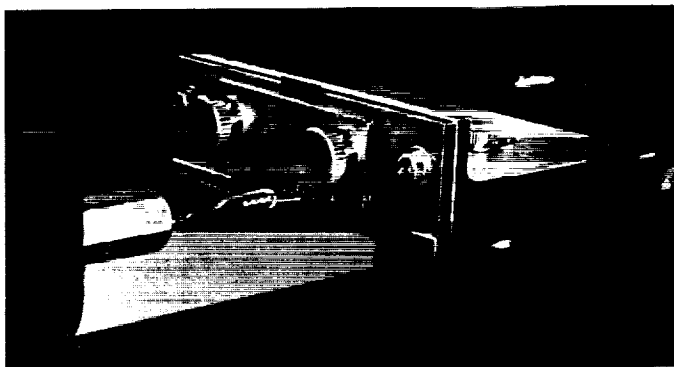
Fig. 11. Solar cell contact strength test, utilizing Hunter tensile tester

The contact strength of the cells was tested by means of a Hunter tensile tester (Fig. 11). Cells were mounted in a specially designed holding fixture attached to the tensile tester, and the P-contact pulled until separation occurred. Preliminary results of this testing are given in Fig. 12. Also shown in Figs. 10 and 12 is the influence of a 6-mil filter glass with 410 mμ cut off and bonded to the cell with LTV 602 adhesive.

The trends indicated by the preliminary data will serve as a basis for determining the direction of future testing. Generally, it was noted that N/P cells appear more capable of surviving long-time temperature storage than do P/N cells.

B. Energy Storage

G. M. Arcand

1. Gravity Effects on Batteries

Experiments employing the cell previously described (Ref. 3) were continued. The only difference in procedure was in the placement of the Viskon separator between the Zn and AgO plates. As mentioned previously, some work was begun in which the separator was placed 5.0 mm from the Zn plate. Further experiments were performed with the separator placed directly against the Zn. The results of these experiments are shown in Fig. 13.
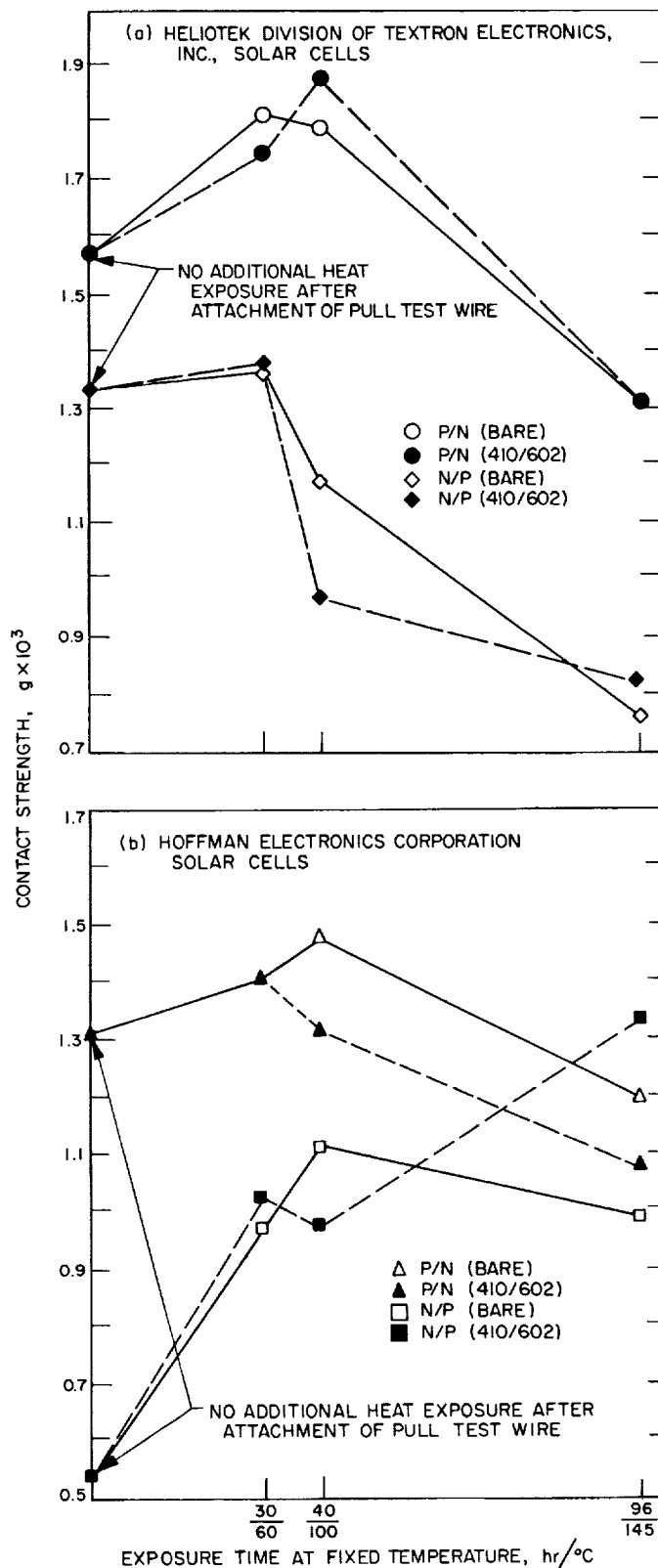
Fig. 12. Contact strength of ohmic strip (light-sensitive side) after exposure to elevated temperatures
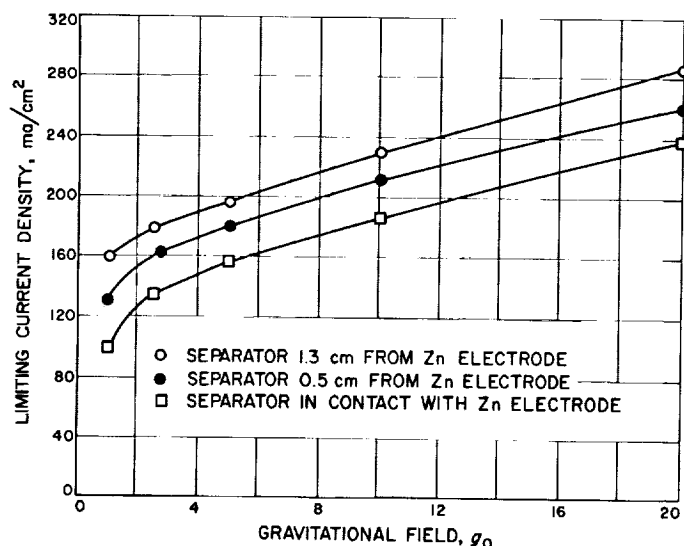
Fig. 13. Effect of gravitational field on cell limiting current

It can be seen from the data that, while the separator does restrict the electrolyte flow somewhat, the relative effect of the gravitational field on convective flow of active material to the electrode remains the same. In the case where the separator is against the Zn, most of the convection may be occurring on the side away from the Zn and diffusion to the plate may occur through the open meshed Viskon.

A cell was assembled with a single layer of sausage casing against the Zn plate in place of the Viskon separator. In this case, the limiting current density was drastically decreased to about 25 ma/cm². Furthermore, approximately the same value was found at both 1.0 and 20.0 g_0, suggesting that only that electrolyte which was trapped between the plate and the separator was available for the reaction and that the current was definitely limited by diffusion rates. Replenishment of the electrolyte by diffusion through the separator was very slow.

A set of experiments is under way for which the cell configuration is the same as that used above. The separator is Viskon placed directly against the Zn, but the electrodes are perpendicular to the gravitational field. Initial indications are that the limiting current density at 1.0 g_0 is less than that occurring when the plates are parallel to the field.

2. Radiation Effects on Battery Behavior

The final report covering the first phase of study on the effects of radiation on NiOOH and Cd electrodes

has been submitted by Atomics International. The results are largely tentative since no actual explanations of observed phenomena are available. The study is being extended for another 6 months.

It was mentioned previously (Ref. 4) that one or both of the electrodes disintegrate to some extent when exposed to high dosages of gamma radiation. The amount of material lost is dependent on the integrated dosage, as is shown in Fig. 14. There is considerable scatter in the data, but the trend is apparent. In addition, the rate of disintegration depends on the dose rate. The average rates of formation of residue were 0.011 and 0.19 mg/hr at dose rates of 8×10^4 and 1.4×10^6 rads/hr, respectively. Again, since the data are scattered, these figures are approximate.

Analysis of the various residues shows that the Cd/Ni ratio varies inversely with the quantity of residue and, hence, the total dosage. The trend is shown in Table 7.
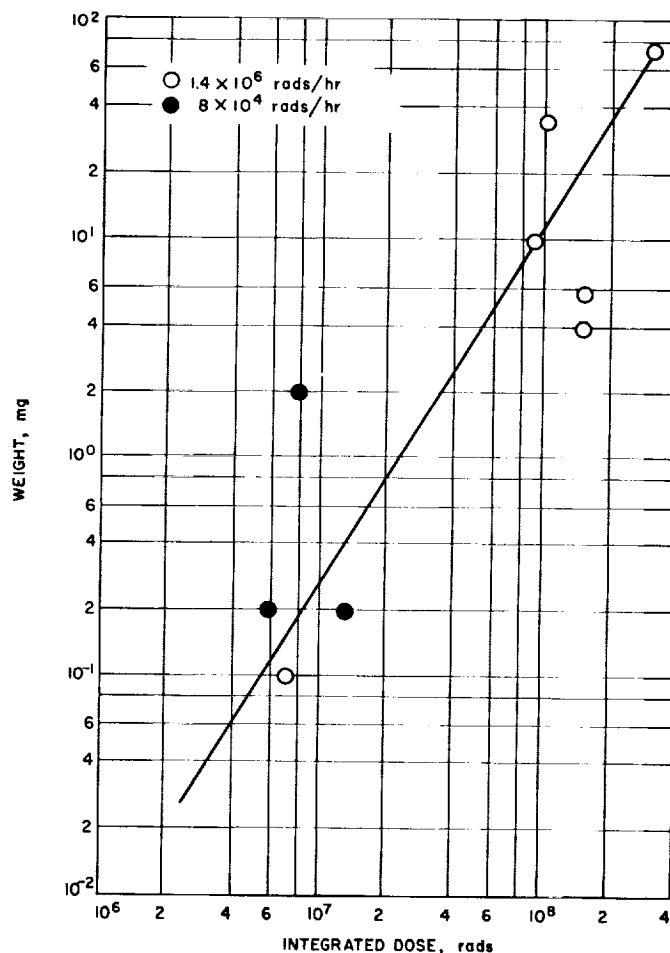
Fig. 14. Weight of residue vs integrated dose

Table 7. Relationship of Cd/Ni ratio of sloughed material to quantity of residue and total dose

Total dose, rads $\times 10^7$	Weight of material, mg	Cd/Ni ratio
1.7	5	10:1
8.7	10.1	7:1
10.0	35.6	6:1
30.0	77	3:1

While both electrodes contain both Ni and Cd, this does not in itself lead to a satisfactory explanation of the change in ratio.

In order to measure gas evolution during irradiation, a stainless-steel container with a quartz liner was fabricated. Electrode dimensions and mounting were the same as those previously described (Ref. 4). The cell was cycled in the gamma field at 1.4×10^6 rads/hr for about 110 hr. The cycle program was a 6-min discharge at 0.2 amp, a 4-min open circuit, a 6-min charge at 0.2 amp, a 4-min open circuit, and then repeat. After the cell was sealed, but before operation began, the pressure dropped to about -1.7 psig, suggesting that oxygen from the air in the cell was reacting with the Cd electrode. After the cell was placed in the Co-60 source, the pressure rose rapidly (probably because of increasing temperatures) and then rose more slowly as the experiment progressed until a positive pressure of about 1.6 psig was reached.

Analysis of the gas indicated that it contained about 8-mol % H_2, 2-mol % O_2, N_2, and traces of CH_4 and CO. Qualitatively, the results can be explained (except for the presence of CH_4 and CO) as a fairly normal radiolysis of an alkaline aqueous solution, as far as is known. The origin of the carbon compounds is unknown.

During the extension of this contract, the radiolysis of concentrated KOH solutions, factors affecting the changes in capacities of the electrode, and more basic electrochemical effects of radiation on the electrodes will be investigated.

References

1. Rouklove, P., "Solar Energy Thermionic Electrical Power Supply Development," SPS 37-25, Vol. IV, pp. 29–35, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

2. Rouklove, P., "SET Concentrator Development," SPS 37-25, Vol. IV, pp. 35, 36, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

3. Arcand, G. M., "Gravity Effects on Batteries," SPS 37-23, Vol. IV, pp. 23, 24, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

4. Arcand, G. M., "Radiation Effects on Battery Behavior," SPS 37-23, Vol. IV, pp. 24, 25, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

III. Guidance and Control Analysis and Integration

A. Development of Advanced Spacecraft Operational Support Equipment

O. E. Linderman

The objectives of this operational-support-equipment (OSE) development task are: (1) to devise more comprehensive OSE testing techniques, (2) to develop and apply margin test techniques for OSE and spacecraft equipment, and (3) to develop other spacecraft tests which result in a more reliably operating spacecraft for the duration of the flight. Subassembly or subsystem tests which will predict an incipient failure of a spacecraft piece part or component are of primary interest.

A digital test circuit which cycles through a short operational routine has been assembled for evaluating various testing techniques. A central computer and sequencer (CC&S) spacecraft subsystem will also be used for this purpose. Possible testing measurements consist of voltage and temperature margin variations or drifts, circuit noise level variations, RF noise radiation, and infrared emission.

1. Digital Test System

The test system (Fig. 1), assembled as a test sample of a typical digital system, uses available NAVCOR (Navigation Computer Corporation) printed circuit cards. The cards include binary counters, shift registers, one-shot delays, binary-coded-decimal–to–decimal converters, and various associated gating circuits. Some inverters, drivers, and delay circuits have been omitted in Fig. 1 for simplification.

The sequence of operation is as follows: With the actuation of the start switch, the 100-kc frequency reference provided by the clock generator (C1) is counted down by binary counters C2 through C6. A division of 2^{23} is achieved with the 23 flip-flops used in the countdown chain. This output is used to turn off the 100-kc reference
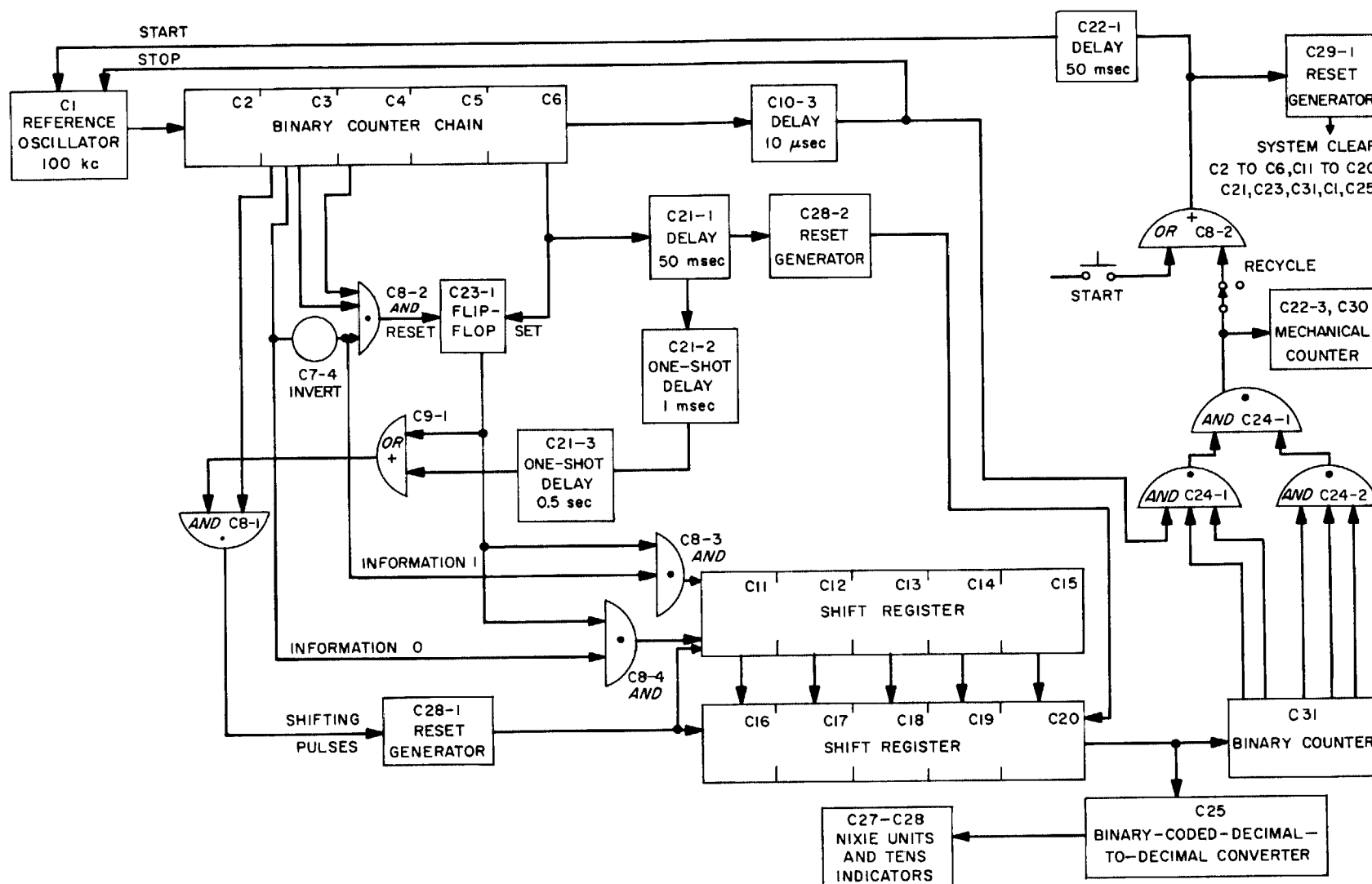
Fig. 1. Digital test system block diagram

at the end of the 83-sec interval. Approximately 21 sec after the start of the sequence, flip-flop C23-1 is set, allowing information being sampled from flip-flop 2 of C3 to be shifted into the C11-C15 shift register. This shifting process is terminated when flip-flop C23-1 is reset after 25 bits of information have been stored in the shift register.

After the 50-msec delay provided by the C21-1 one shot, the information stored in C11-C15 is transferred in parallel to the C16-C20 shift register. Then, 1 msec later (the C21-2 delay), the C11-C15 shift register is cleared and the information in the C16-C20 register is counted and decoded by C25, C26, and C27 for Nixie display. This display is the decimal equivalent (number 13) of the total number of binary *ones* counted by C25. The above shifting process is repeated a second time approximately 40 sec later, with the result that the Nixie indicators show a total count of decimal 26.

Binary counter C31 also counts the number of binary *ones* shifted from the C16-C20 shift register. If the count is correct, the C24-4 *and* circuit allows the 100-kc frequency reference to start again for the next cycle, provided the system is in the recycle mode. On the successful completion of a normal sequence (obtaining correct count from C31), a mechanical counter is advanced by 1. Thus, the sequence is repeated continually (and the mechanical counter advanced by 1 each cycle) by operating the system in the recycle mode. The system stops and will not recycle if there is a failure or if the sequence does not present the proper count to C31.

The rack mounting of the circuit cards is shown in Fig. 2. The means by which a circuit card is mounted for taking infrared photos and the cabling to place the circuit card tested in the environmental chamber can be seen.

2. Status of Testing Techniques

Various tests completed on the test system indicate that more precise measurements must be made and testing techniques must be refined.

a. Voltage and temperature margins. The circuit cards have been individually tested for positive and negative voltage supply variation of 25 to 30% above and below the nominal at ambient temperatures. Only the counter cards failed to function at these extremes. The counter cards were then checked over this same voltage range at various temperatures from 0 to 150°F. This informa-

Fig. 2. Digital test system rack

tion has been plotted in a "shmoo" diagram for reference condition, and repeatability measurements are being made. More complete factory test information is being sought on other cards in the test system to obtain more meaningful margin test information for these cards.

b. Circuit noise level. Test circuits are being studied and design information is being sought to determine noise susceptibility. Present measurements have not indicated that circuits are susceptible to the injection of noise of several-volts magnitude, although spikes in the counter chain do give an erratic count.

c. RF noise radiation. More details of this testing technique will be obtained from reports by its proponent.

d. Infrared emission. Infrared photos have been taken of the digital test system cards and each of the CC&S modules. Circuit card 15, one of the shift registers, is shown in Fig. 3. Ambient (or gray) background is about 74°F, while the white of the gray scale is about 32°F above ambient or 106°F. Fig. 4 shows the *Ranger* CC&S power transformer–rectifier module. The ambient temperature at the center top edge of the module is 94°F,
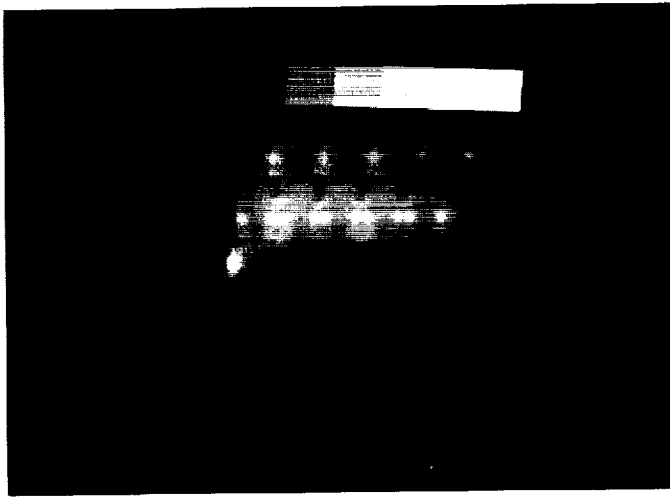
Fig. 3. Shift register, circuit card 15

Fig. 4. Transformer—rectifier module of the *Ranger* central computer and sequencer

while the two large white areas (power diodes) are at 126°F.

Infrared photos readily show various modes of operation, such as: (1) the duration that electronic switches are closed, or (2) which register contains a 1 or 0. Slight variations or drifts in the many components are more difficult to access. Voltage measurements in the infrared detector amplifier loop of several hot components on the circuit cards have been taken on a brush recorder to determine day-to-day repeatability. Measurements of a zener diode have been found to vary by 15%. An attempt is being made to resolve these variations in temperature readings in terms of changes in circuit conditions, the infrared measurement equipment, or the operation technique.

Improvement of the method of mounting the infrared radiometer and subject card being measured appears to be desirable. Investigation of the possibility of mounting the radiometer on a lathe or milling machine is being delayed until a new and improved infrared camera is received.

Infrared radiation measurements of a cold soldered joint were not noticeably different from those of a properly made joint. These measurements were taken for various power dissipations in resistors connected to the joints. An attempt will be made to measure the heat gradient through good and bad joints where one of the conductors to the joint is heated and the variation is determined in heat flow through different joints.

IV. Guidance and Control Research

A. Magnetics Research

K. Kuwahara, F. B. Humphrey, and M. Takahashi

1. Thickness Dependence of the Magneto-resistance Effect in Thin Permalloy Films,
K. Kuwahara

The magnetoresistance effect is the change in resistance observed when applying a magnetic field to a sample. It arises from the interaction between conduction carriers and magnetic fields applied and/or induced; therefore, any material having conduction carriers exhibits the magnetoresistance effect. The ferromagnetic metals, however, show an extraordinarily large magnetoresistance effect at low fields compared to the normal metals. This enhancement may be attributable to: (1) the large internal field or spontaneous magnetization, which is easily varied in direction by applying relatively low fields; and (2) the strong coupling between conduction electrons and magnetic electrons (Ref. 1).

A phenomenological relation between the magnetoresistance ΔR and the magnetization M is easily obtained.

Accordingly, ΔR is available as a tool to investigate the magnetic process in ferromagnetic metals (Ref. 2). Especially in the case of thin films, ΔR measurements are excellent, because the effect is large enough that measurements can be made with accuracy even if the film is very thin. In this discussion the general aspects of ΔR in thin permalloy films and preliminary results of the thickness dependence of the magnetoresistance effect will be presented.

a. Magnetoresistance effect in permalloy films. Consider the relationship between the resistance and the magnetization. In Fig. 1, i and M are the current for measuring resistance and the saturation magnetization, respectively. Both are in the plane of the film. The resistance R, expressed as an expansion of the angle θ between i and M, is:

$$R = R_{\shortparallel} (1 + a \sin^2 \theta + b \sin^4 \theta + \cdots), \qquad (1)$$

where $R_{\shortparallel}$ is the resistance when $\theta = 0$, and $a, b, \cdots$ are coefficients. The coefficients of the terms of odd power
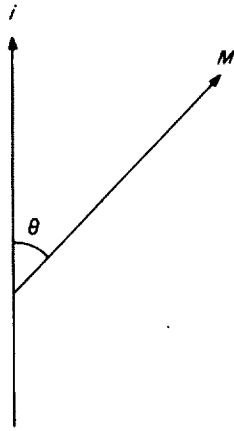
Fig. 1. Electric current *i* for measuring the magneto-resistance and magnetization *M* in the plane of the film

of sin θ are zero because of symmetry. Experimentally it is confirmed that R can be expressed as:

$$R = R_{||}(1 + a\sin^2\theta). \qquad (2)$$

This equation can be rewritten using $R_\perp$ as:

$$\frac{\Delta R}{R_\perp} = \frac{R_{||} - R_\perp}{R_\perp}\cos^2\theta, \qquad (3)$$

where $R_\perp$ is the resistance when $\theta = \pi/2$ and ΔR is the change of resistance from the reference state (i.e., when $\theta = \pi/2$). Fortunately, in permalloy films near the composition 80-20 Ni-Fe, the fractional change in resistance is fairly large. Fig. 2 is an example of the relationship expressed in Eq. (2) and of this large change for a typical permalloy film. The term $(R_{||} - R_\perp)/R_\perp$ is a coefficient which determines the magnetoresistive character of the film and depends on the composition and temperature (e.g., Ref. 3). As will be seen later, this term also depends upon the thickness.

With magnetoresistance measurements it is possible to obtain H_c, H_K (Ref. 4), and dispersion. The mechanism of magnetization flux change, both quasi-static and dynamic, can be inferred by observing the ΔR curve during the process. Only the method for obtaining the dispersion of M will be considered here. When the film is saturated along its easy direction with a field greater than about 10 H_c, M is aligned to that direction. After removing the field, M will assume some equilibrium value according to the dispersion of H_K in the film; hence, the resistance will change. If the direction of
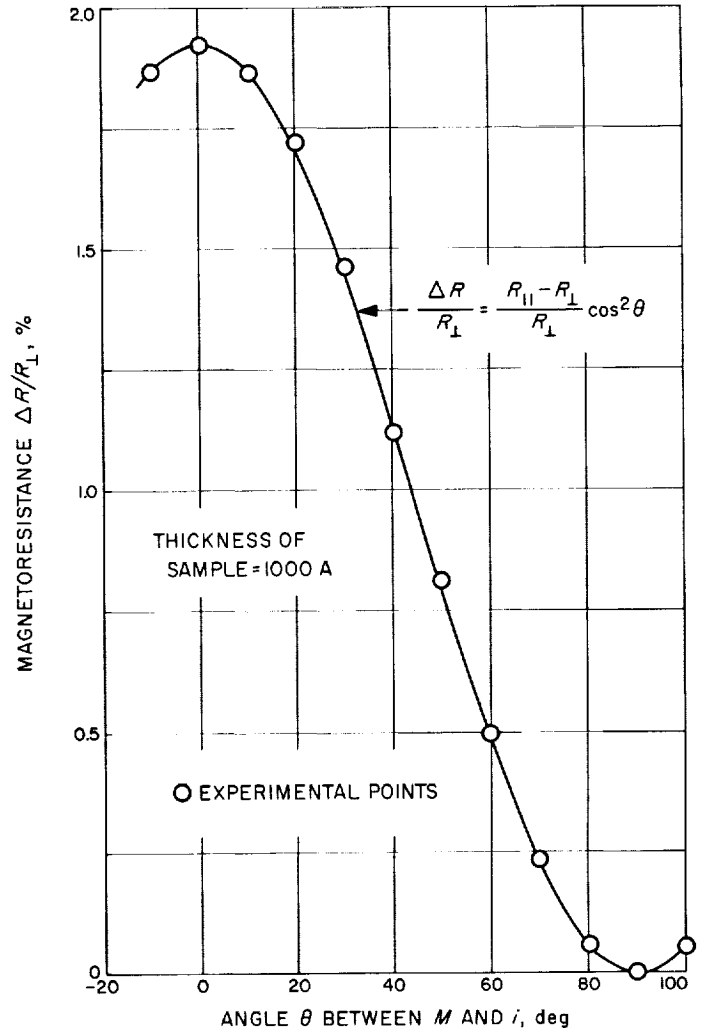
Fig. 2. An example of resistance change in thin films of 83-17 Ni-Fe

the current is parallel to the easy direction, the resistance is changed from $R_{||}$ to R_r. The average dispersion α can be obtained approximately as:

$$\frac{1}{2\alpha}\left(\alpha - \frac{1}{4}\sin 4\alpha\right) = \frac{R_{||} - R_r}{R_{||} - R_\perp}. \qquad (4)$$

b. Thickness dependence of ΔR. Thin films of permalloy were produced by vacuum evaporation from a melt of 83-17 Ni-Fe onto a hot (300°C) glass substrate with a deposition rate of 1000 A/min, as was done previously (Ref. 5). The films for this experiment were 1×1 cm², with electrodes of evaporated gold at both ends. The current was parallel to the easy axis for all the experiments. A field of about 200 oe was applied in the plane of the film, parallel to and then perpendicular to the easy

axis. The potential drop between the electrodes was measured as $V_\parallel$ and $V_\perp$, respectively. The magnetoresistance coefficient $(R_\parallel - R_\perp)/R_\perp$ is given by $(V_\parallel - V_\perp)/V_\perp$. The value of this coefficient is plotted as a function of the magnetic thickness in Fig. 3. Although the points are considerably scattered, the coefficient clearly decreases with decreasing thickness below about 600 A.

Resistivity of the films can be calculated using their size and magnetic thickness. The thickness dependence of resistivity is shown in Fig. 4. In spite of the scattering, it appears that the resistivity increases below 300 A.
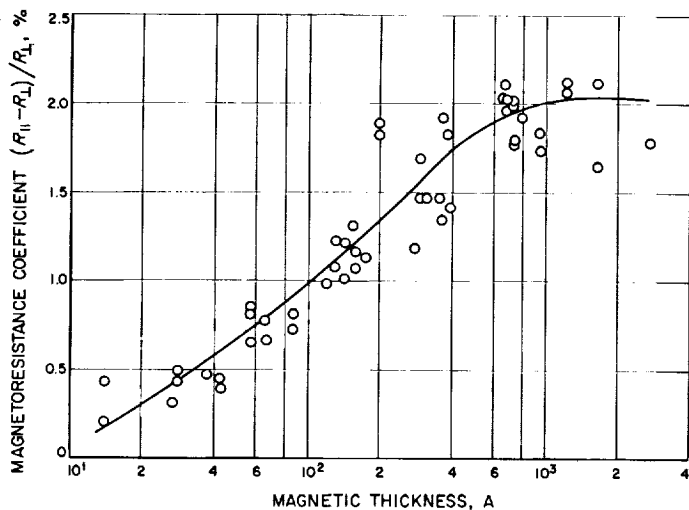
Fig. 3. Dependence of the magnetoresistance coefficient upon the magnetic thickness of the film

It is interesting to consider the difference in resistivity $(\rho_\parallel - \rho_\perp)$ when M is parallel and perpendicular to i in the same way as the magnetoresistance coefficient $(R_\parallel - R_\perp)/R_\perp$ was considered. As can be seen in Fig. 5, the scattering of points is less than that for $(R_\parallel - R_\perp)/R_\perp$. The main feature in Fig. 5, however, is still the same as that in Fig. 3. The change in resistivity $(\rho_\parallel - \rho_\perp)$ definitely decreases with decreasing thickness from about 200 A.

As observed in the experiments of microwave absorption (Ref. 6), Hall effect (Ref. 7), and torque measurement (Ref. 8), the saturation magnetization of permalloy films having the composition near 80-20 Ni-Fe seems to decrease with decreasing thickness. The rapid decrease in M occurs for films thinner than about 100 A. In the present experiment, either $(R_\parallel - R_\perp)/R_\perp$ or $(\rho_\parallel - \rho_\perp)$ also decreases with decreasing thickness and seems to vanish at a very small thickness. Though the intrinsic relationship between the magnetoresistance and the saturation magnetization is not well-known, it is suspected that the magnetoresistance is closely connected to the magnitude of M (Ref. 1). Then, such a decrease of magnetoresistance effect as observed here is also considered a result of the decrease in M, even though the decrease in the magnetoresistance begins with thicker films. As mentioned above, a torque measurement (Ref. 8) was made on the films having the same composition which were produced in the same apparatus as those of the present experiment. The magnetic thickness can then be referred to the true thickness. The thickness dependence of $(\rho_\parallel - \rho_\perp)$ after

Fig. 4. Dependence of resistivity upon the magnetic thickness of the film
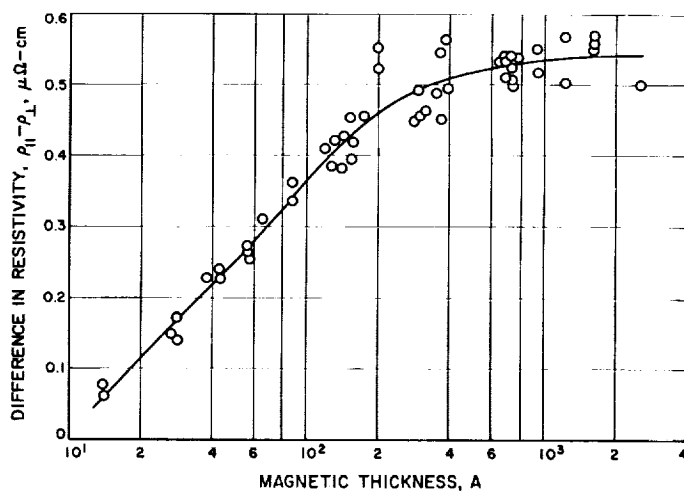
Fig. 5. Resistivity difference when M is parallel and perpendicular to i as a function of the magnetic thickness of the film
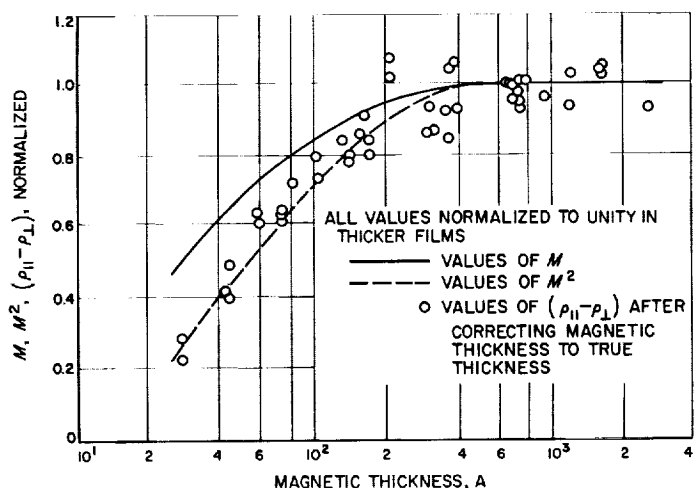
Fig. 6. Comparison of the thickness dependence of
M, M^2, and $(\rho_{\parallel} - \rho_{\perp})$

this correction was made is shown in Fig. 6. Also shown is the thickness dependence of M and M^2, using the value of M obtained by the torque measurement (Ref. 8). (All values are normalized.) It can be seen that the decrease in $(\rho_{\parallel} - \rho_{\perp})$ is proportional to the decrease of M^2 rather than that of M.

c. Thickness dependence of the angle dispersion of M. As mentioned previously, the angle dispersion of M caused by dispersed H_K can be obtained by measuring ΔR. In Fig. 7, the average dispersion α is plotted as a function of thickness. For films near 1000 A, the dispersion is small and only a little scattering is observed in the results. The dispersion seems to increase with
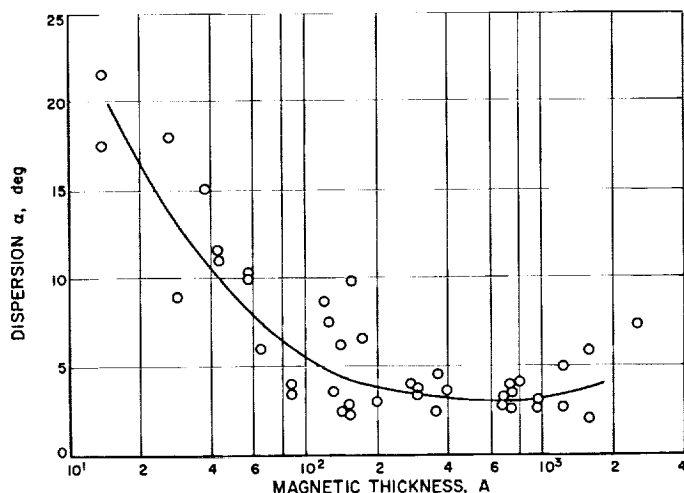
Fig. 7. Dependence of the dispersion α upon the magnetic thickness of the film

decreasing thickness and becomes significant at about 200 A. The large dispersion observed in extremely thin films is presumably related to the island structure in such thin films.

2. Induced Anisotropy in Iron–Cobalt Films,

F. B. Humphrey and M. Takahashi[1]

Almost all investigations on induced magnetic anisotropy in thin films have been done with various combinations of Ni and Fe. Slight additions of Co have been made in an attempt to increase the anisotropy independent of the coercive force. No systematic study of the Fe-Co system has been made, even though the metallurgy of the bulk material is complicated but fairly well-known. Actually, magnetic measurements with bulk material through the entire composition range are very difficult because of the physical properties of the Fe-Co alloys. Since evaporated thin films do not present this difficulty, a comprehensive study of Fe-Co thin films through the entire composition range is being made. Particular emphasis is being placed on the induced anisotropy and rotational hysteresis loss. The saturation magnetization, remanent magnetization, and coercive force are also being observed. Only the preliminary results on the induced uniaxial anisotropy will be reported here.

The films were made by vacuum evaporation, at a pressure of about 2×10^{-7} torr, from a melt of the vacuum cast alloy. For a melt temperature of about 1400°C, yielding an evaporation rate of 1000 to 1500 A/min at 19 cm, the chemical composition of the resultant film as a function of the melt composition can be seen in Fig. 8. The composition of the melt was determined by standard wet chemical analysis, and the composition of the films was measured by X-ray fluorescence (Ref. 9). Generally, two 1-cm D films were made at a time with the glass substrate at room temperature, and two films were made with the substrate at 300°C. The film thicknesses, also measured by X-ray fluorescence, varied from 500 to 1500 A.

All films exhibited a uniaxial magnetic anisotropy in the plane of the film. The direction of the easy axis of this anisotropy was aligned parallel to an applied magnetic field in the plane of the substrate during evaporation. The magnitude of the anisotropy was measured at room temperature using a torquemeter and the method previously described in Refs. 10 and 11.
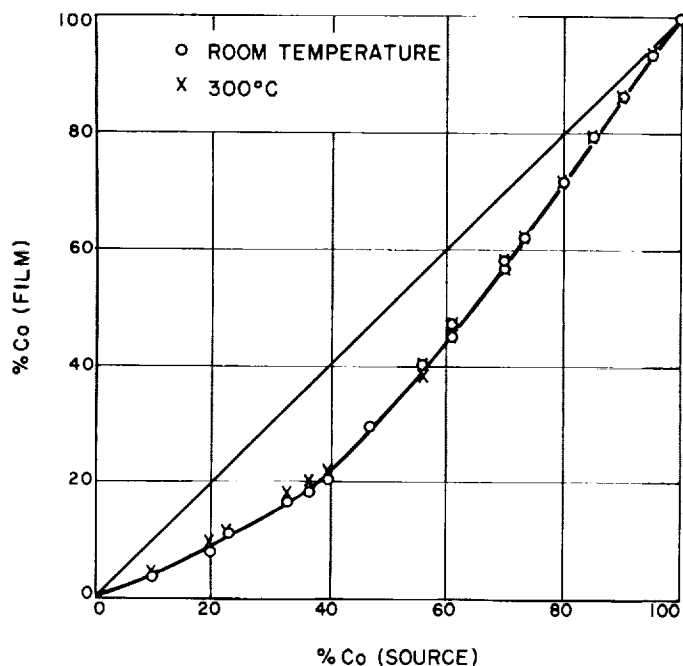
Fig. 8. Film composition as a function of melt composition for vacuum-evaporated Fe-Co
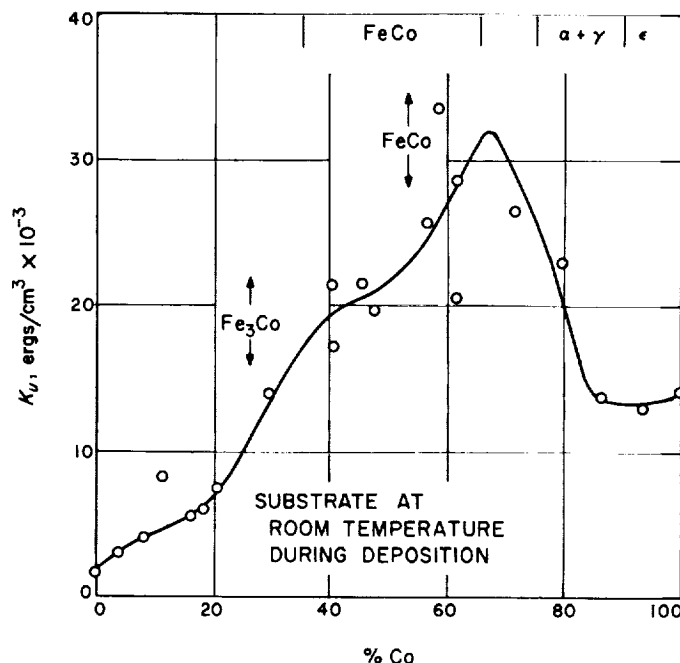
Fig. 9. Induced uniaxial anisotropy in Fe-Co thin films made with the substrate at room temperature during deposition

Torque curves were taken at various fields up to about 500 oe. The value of anisotropy was obtained by the linear extrapolation to infinite field on a plot of the peak value of uniaxial torque as a function of $1/H$, as is generally done for bulk materials (Ref. 12, p. 566). The magnitude of the uniaxial anisotropy constant, K_u, is shown as a function of composition in Fig. 9 for films made with the substrate at room temperature during deposition, and in Fig. 10 for films made with the substrate at 300°C during deposition. The expected phase of bulk material is indicated at the top in Figs. 9 and 10.

The first striking feature of the data is the magnitude of the induced anisotropy. Fig. 11 shows anisotropy of the Fe-Ni system for comparison (Ref. 13). The magnitude of the anisotropy constant in Fe-Co is nearly an order-of-magnitude greater than in Ni-Fe films. Also, in the composition range 30 to 70% Co, the anisotropy for a 300°C substrate is slightly greater than that for a film made at room temperature. This inversion is quite unexpected, since all other data are similar to those in Fig. 11, where the anisotropy for films made at room temperature is greater than that for films made at elevated temperatures.

One possible mechanism for magnetic anisotropy in this system is the Néel–Taniguchi directional order of atom pairs (Refs. 14 and 15). The induced anisotropy
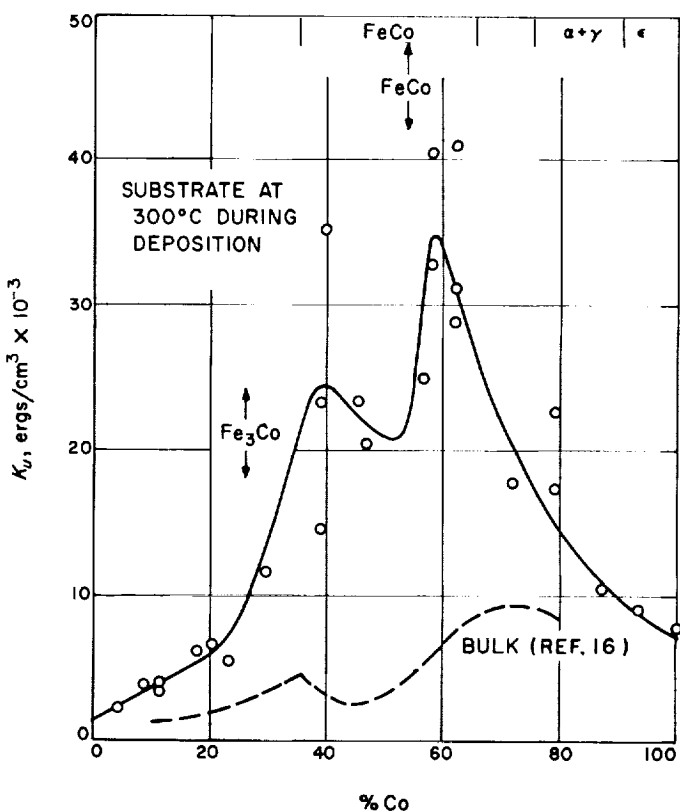
Fig. 10. Induced uniaxial anisotropy in Fe-Co thin films made with the substrate at 300°C during deposition
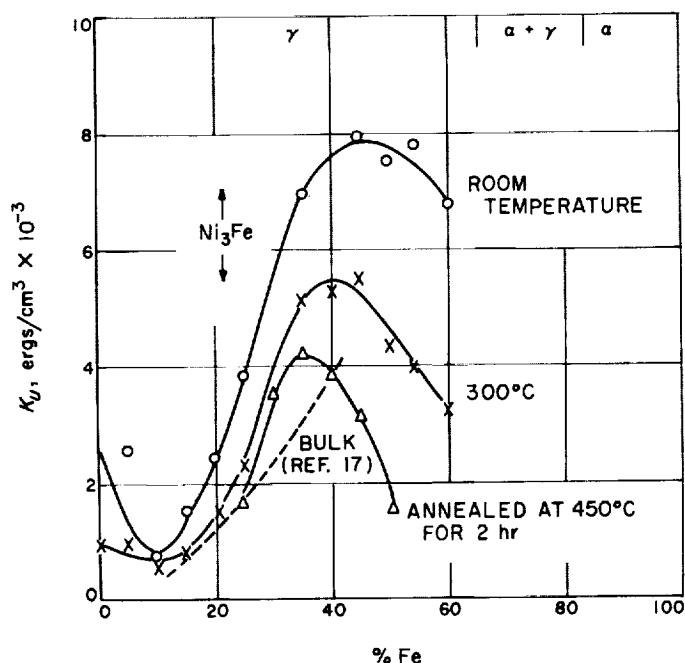
Fig. 11. Induced uniaxial anisotropy in Fe-Ni thin films

for bulk Fe-Co (Ref. 16) and Fe-Ni (Ref. 17) is given in Figs. 10 and 11, respectively. In both cases it is generally accepted that the anisotropy is caused by pair orientation. For this type of anisotropy, there is a significant decrease near the composition Ni_3Fe, the ordered structure in the γ (i.e., fcc)[2] phase, and near FeCo, the ordered structure in the α (i.e., bcc) phase. As can be seen in Figs. 9 and 10, there seem to be inflections in the range of the ordered structure composition Fe_3Co and FeCo. This observation leads to the conclusion that pair orientation is one likely mechanism for this anisotropy. It should be noted, however, that the shape of the curves in Figs. 9 and 10 is similar to the shape of the curves for magneto-striction as a function of composition (Ref. 12, p. 664). Perhaps it is necessary to consider magnetostriction as indicated in the case of Ni-Fe (Fig. 11), where the peak in anisotropy for a room-temperature substrate is at the proper concentration to be explained by pair orientation, but where the shift in the peak as the substrate temperature is increased cannot be attributed to pair orientation. Here, also, magnetostriction is suggested.

Structural imperfections in the form of stacking faults, dislocations, and vacancies, coupled with impurities such as oxygen (Ref. 18), might form the basis for induced magnetic anisotropy. The structure of the Fe-Co films has not, as yet, been determined, but a reasonable guess

can be made by considering the structure data for bulk material (Ref. 19). Since there is a $\gamma \rightarrow \varepsilon$ transition (i.e., fcc $\rightarrow$ hcp) about 400°C, pure Co will probably contain many structural imperfections. Also, many structural imperfections can be expected between 75 and 90% Co since this range is a mixture $(\alpha + \gamma)$ phase at room temperature. A single phase (α) is expected as the Co concentration is decreased below the mixture range, so 60% Co should be well into the single-phase region. If the mechanism of anisotropy involves crystalline imperfections, it would be expected that the magnitude of the anisotropy would depend upon the density of imperfections. It can be seen that the induced anisotropy at various compositions in Figs. 9 and 10 does not follow the expected imperfection density arguments, leading to the conclusion that crystalline imperfections, as such, are probably not important in this system.

The preliminary results on the Fe-Co system look very promising. Work will continue and will be reported in subsequent issues of the SPS. It is expected that other magnetic characteristics will be considered, along with other possible mechanisms for the observed anisotropy.

B. Optical Physics Research

J. M. Weingart and A. R. Johnston

1. Electro-optic Coefficients of Single Crystals of Barium Titanate

Meyerhoffer (Ref. 20) and Hornig[3] have shown that the refractive properties of single-crystal barium titanate in the ferroelectric tetragonal state and the paraelectric cubic state can be altered considerably by an external electric field. The purpose of the work reported here was to accurately measure these effects as a function of temperature and field strength in the tetragonal phase and to compute the corresponding electro-optic tensor components.

In general, the tensor components depend both linearly (Pockels effect) and quadratically (Kerr effect) on the external electric field, but the quadratic effect does not become significant until approximately 100°C (Ref. 20). Throughout most of the ferroelectric range (5 to 120°C), the effect is essentially linear.

---

[2]The face-centered cubic, body-centered cubic, and hexagonal close-packed phases are signified by fcc, bcc, and hcp, respectively, in this discussion.

[3]Hornig, A., *Electro-optics and Paramagnetic Resonances in Barium Titanate Single Crystals*, Ph.D. Thesis, Stanford University, 1955 (Unpublished).

The mathematics necessary to relate the experimentally observed quantities to the theoretical parameters will be presented first. The notation adopted by Billings (Ref. 21) and others will be used. The refractive properties of crystals are completely specified by the index tensor a_{ij}. Since $a_{ij} = a_{ji}$, the index tensor can be represented by a convenient geometric construction, the "index ellipsoid," defined as the surface:

$$a_{ij}x_ix_j = 1. \tag{1}$$

This geometric representation is useful since the principal directions and associated indices for a plane wave propagating through the crystal are represented by the orientation and lengths of the semimajor and semiminor axes of the elliptical intersection of the index ellipsoid and a plane through the center of the ellipsoid normal to the direction of propagation. The meaning of the index ellipsoid is treated in detail in many texts, e.g., Ref. 22.

The linear electro-optic effect is defined as the case when the a_{ij}'s are linear functions of the external field. It can be visualized in the most general case as a change in orientation and dimensions of the index ellipsoid. In certain cases, the dimensions of the ellipsoid may depend quadratically on the field when the a_{ij}'s change linearly with the applied field. This should not be confused with the Kerr effect, where the tensor components a_{ij} depend quadratically on the external electric field. The dependence of the refractive properties of the medium on applied electric field is given by the changes in orientation and dimensions of the index ellipsoid as a function of the external field **E**. This can be reduced to a standard eigenvalue problem (Ref. 23). The field-dependent ellipsoid is:

$$(a_{ij} + \Delta a_{ij}(\mathbf{E}))x_ix_j = 1. \tag{2}$$

The quadratic form, Eq. (2), can be represented by the matrix $[a'] = [a] + [\Delta a]$. The eigenvectors $\mathbf{X}_r$ and the eigenvalues λ_r for the matrix equation

$$([a'] - \lambda_r[I])\mathbf{X}_r = 0 \tag{3}$$

are then the perturbed principal directions and inverse squares of the associated principal indices, respectively; i.e.,

$$\lambda_i = n_i^{-2}. \tag{4}$$

In tensor notation the linear electro-optic effect is written

$$\Delta a_{ij}(\mathbf{E}) = r_{ijk}E_k, \tag{5}$$

where r_{ijk} is defined as the linear electro-optic tensor. Since the equation is symmetric in the indices i and j, these indices are usually contracted in the following way (Ref. 22):

Tensor notation	11	22	33	23	13	12
Matrix notation	1	2	3	4	5	6

Eq. (5) is written in matrix form as follows:

$$\Delta a_i = r_{ik}E_k, \tag{6}$$

where i may vary from 1 to 6 and k from 1 to 3.

The coordinate system chosen is the conventional (orthogonal) crystal system (Ref. 22), which is the principal axis system for the field-free ellipsoid in barium titanate. The form of the matrix r_{ik} for the 20 non-centro symmetric crystal classes which can exhibit a linear electro-optic effect is given in Ref. 22. Below the fundamental mechanical resonant frequency of a crystal, the above matrix includes a piezoelectric-photoelastic contribution which, in barium titanate, is of the same form as that in the strain-free case (Ref. 24).

For barium titanate in the tetragonal (4mm) phase, with the z-direction along the unperturbed polar axis, the matrix equation for Δa_i is:

$$
\begin{bmatrix} \Delta a_1 \\ \Delta a_2 \\ \Delta a_3 \\ \Delta a_4 \\ \Delta a_5 \\ \Delta a_6 \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & r_{13} \\ \cdot & \cdot & r_{13} \\ \cdot & \cdot & r_{33} \\ \cdot & r_{42} & \cdot \\ r_{42} & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \cdot \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}
$$

The dots correspond to the matrix elements which are zero. The equation of the index ellipsoid in the presence of a field $\mathbf{E} = (E_x, E_y, E_z)$ is then, from Eq. (2),

$$(a_{11} + r_{13}E_z)(x^2 + y^2) + (a_{33} + r_{33}E_z)z^2$$

$$+ 2r_{42}(E_xxz + E_yyz) = 1. \tag{7}$$

The components r_{13} and r_{33} describe the effect due to a c-directed field, and the r_{42}'s describe the effect due to an a-directed field. Note that

$$a_{11} = \frac{1}{n_a^2},$$

$$a_{33} = \frac{1}{n_c^2},$$

where n_a is the index of refraction for a light wave having its E vector along an a-axis, i.e., in the x or y direction; similarly, n_c applies when the E vector is parallel to the c-axis.

By symmetry it is clear that, for a c-directed field, the principal directions will not change and the crystal will remain uniaxial. This is verified by noting that the matrix $[a']$ remains diagonal for a c-directed field:

$$\begin{bmatrix} (a_{11} + r_{13}E_c) & 0 & 0 \\ 0 & (a_{11} + r_{13}E_c) & 0 \\ 0 & 0 & (a_{33} + r_{33}E_c) \end{bmatrix} \quad (8)$$

Assuming $r_{13}E_c \ll a_{11}$ and $r_{33}E_c \ll a_{33}$, the perturbed indices $n_c' = (a_{33}')^{1/2}$ and $n_a' = (a_{11}')^{1/2}$ are related to the zero-field indices by:

$$n_c' = n_c + \tfrac{1}{2} r_{33} n_c^3 E_c, \quad (9a)$$

$$n_a' = n_a - \tfrac{1}{2} r_{13} n_a^3 E_c. \quad (9b)$$

The validity of this assumption is verified later in this discussion. There is no practical way to separate the two coefficients r_{13} and r_{33} since the only easily measured quantity related to them is the induced birefringence, which depends on their difference. For convenience the term

$$r_{33}' = r_{33} - (n_a/n_c)^3 r_{13}$$

is introduced, so that the field-induced birefringence can be written as:

$$\Delta n = \Delta (n_c' - n_a') = \tfrac{1}{2} r_{33}' n_c^3 E_c. \quad (10)$$

This is equivalent to defining $r_{13} = 0$.

Consider now an a-directed field, where $E_z = E_y = 0$ and $E_x = E_a$. The eigenvalue problem is given by:

$$\begin{bmatrix} a_{11} - \lambda_r & 0 & r_{42}E_a \\ 0 & a_{11} - \lambda_r & 0 \\ r_{42}E_a & 0 & a_{33} - \lambda_r \end{bmatrix} \cdot X_r. \quad (11)$$

The new principal directions are found by solving for the eigenvectors X_r in Eq. (11), again assuming $r_{42}E_a$ to be small.

$$X_1 = \left(1, 0, \frac{r_{42}E_a}{a_{11} - a_{33}}\right)$$

$$X_2 = (0, 1, 0) \quad (12)$$

$$X_3 = \left(\frac{-r_{42}E_a}{a_{11} - a_{33}}, 0, 1\right)$$

Thus, a field in the x-direction results in a first-order rotation of the principal directions in the plane containing the applied field and polar axis. Designating the angle of rotation by θ,

$$\theta \simeq \frac{r_{42}E_a}{a_{11} - a_{33}} = \left(\frac{n_a^2 n_c^2}{n_c^2 - n_a^2}\right) r_{42}E_a. \quad (13)$$

a. Experimental. In addition to determining the dependence on temperature, the induced changes in optical behavior were measured at room temperature as a function of strength and frequency of the applied field. All measurements were made using the optical bench and polarimeter described in Ref. 25. For measurements as a function of temperature, the mounted crystals were placed in an oven designed to operate to 150°C. The oven (Fig. 12) consists of a 3-in. D brass tube wrapped with Nichrome wire and insulated by an Epiglass sleeve. The endplates are brass, with Epiglass plates again serving as thermal insulation. This brass tube contains a fixture which holds the sample mounts. The clamps which hold the mount to the fixture also serve to make electrical contact and are connected to two pieces of Teflon-insulated coaxial cable. The entire oven assembly fits into a mounting which rotates about an axis passing through the center of the crystal sample. A vernier dial permits the reading of this rotation to 1 min of arc.

Two copper constantan thermocouples were used for temperature measurements: one being attached to the crystal and the other measuring ambient air temperature inside the oven. Optical measurements were made when the two thermocouples registered within a few tenths

Fig. 12. Sample oven

of a °C of each other to ensure that the measurements were being made under conditions of thermal equilibrium.

Measurements were made on two types of single-crystal barium titanate: r'_{33} with Remeika-type flux-grown crystals (Ref. 26), which were obtained from the Harshaw Chemical Company; and r_{42} with crystals grown by a radically different method from a titania melt, which were supplied by Professor Arthur Linz of the Laboratory for Insulation Research, Massachusetts Institute of Technology. The Linz crystals were made available in the form of slices from a boule, from which rectangular samples 1 mm thick $\times$ several mm on an edge could be cut, with the c-axis in the plane of the sample and parallel to one edge.

Thin plates of barium titanate which contain the polar c-axis in the plane of the plate are conventionally called a-plates. The poling of flux-grown crystals was discussed in detail in Ref. 27. Flux-grown triangular wings 3 to 8 mm on an edge were placed in concentrated H_3PO_4, heated to 150°C, and etched for 2 min. After etching and cooling, the crystals were rinsed several times with distilled water and cleaved along the natural (100) directions, using a watchmaker's staking tool with a punch ground in the shape of a sharp chisel. After electroding and mounting, the crystals were poled by heating above the Curie point (120°C) and cooling through the transition temperature while maintaining a dc field of 2 kv/cm in the plane of the plate. Poling of the Linz samples to remove antiparallel domains was achieved by temporarily applying a dc field of 1500 v/cm across the crystal in the c-direction at room temperature, using distilled water for electrodes.
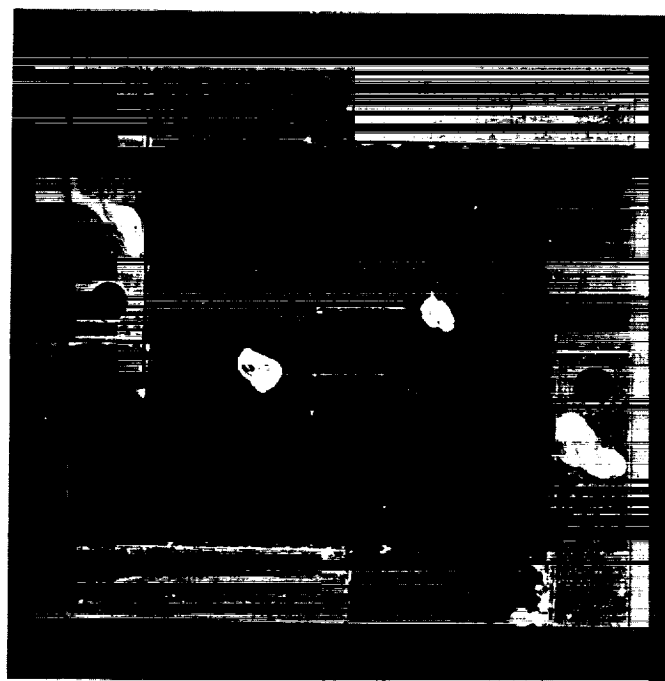
Fig. 13. Typical mounted sample

Both types of crystals were electroded on opposite edges with Aquadag (colloidal graphite); 1-mil gold wires served as lead-ins. The crystals were mounted on 1-in. squares of Epiglass (epoxy impregnated fiberglass). Strips of copper bonded to the squares near opposite edges served as terminals (Fig. 13). Samples A (3.12 $\times$ 3.20 $\times$ 0.174 mm) and B (2.70 $\times$ 2.70 $\times$ 0.180 mm) were flux-grown a-plates, electroded to provide an electric field in the c-direction, and these were used to determine r'_{33}. Crystal C (1.10 $\times$ 4.12 $\times$ 3.68 mm) was a Linz sample, an a-plate electroded to provide a field in an a-direction, and this was used to determine r_{42}. Unfortunately, time limitations prevented the comparison of the two types of crystals.

b. Applied field in the c-direction. The experimental configuration for measuring r'_{33} is shown in Fig. 14. The crossed polarizer and analyzer axes make an angle of 45 deg with the polar c-axis of the sample. The quarter-wave plate axes are parallel to the crystal axes (Senarmont configuration; see Ref. 28). This configuration is useful for measuring retardation, provided the orientation of the principal directions is field-independent.

If the light entering the crystal were perfectly monochromatic, it would emerge from the crystal with a unique ellipticity and would thus emerge from the quarter-wave plate being linearly polarized and light-oriented at an angle ϕ with respect to the quarter-wave plate axes,
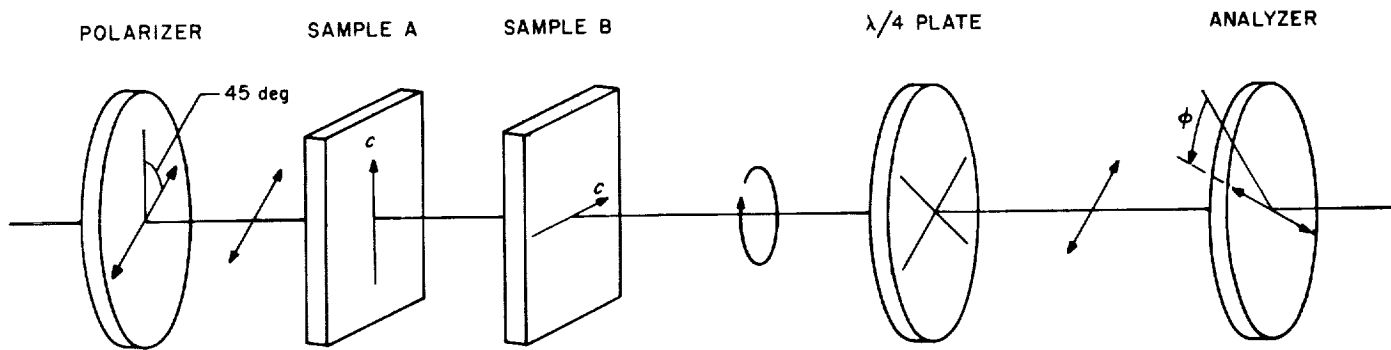
Fig. 14. Experimental setup for measurement of r'_{33}

where ϕ depends on λ. An additional field-induced retardation could be measured by measuring the change in ϕ, regarding the initial orientation as a dc optical bias. In the Senarmont configuration $\Delta\phi = \frac{1}{2}\delta$, where δ is the retardation measured as a phase angle.

However, due to the 20-A bandwidth of the source plus interference filter, the wavelength-dependent orientation of the linearly polarized light incident on the analyzer is spread over about 1 rad for a crystal 0.2 mm thick, making accurate measurement impossible. To reduce this angular spread, another crystal, matched as closely as possible in thickness to the first, was mounted in back of the first crystal with the polar axes of the two crystals oriented at right angles (Fig. 14). The resulting spontaneous birefringence then depends on the differential thickness, which was reduced to 0.01 mm. This arrangement proved to be satisfactory.

Both of the electro-optic coefficients were found to be essentially frequency-independent up to 100 kc. The fundamental piezoelectric resonance is about 750 kc with the size of samples used. However, with dc fields, space charge can build up in the crystals, neutralizing to some extent the effect of the external field. This effect in barium titanate (Ref. 29) was observed to varying degrees in the crystals used in this experiment. At room temperature the time constant for this buildup was observed to be about 8 sec for Sample A. In addition, the size of the effect can vary from point to point in individual crystals, as well as from crystal to crystal. In order to minimize this space-charge effect, measurements were made with an ac (1-kc) applied field.

Fig. 15 shows the optical retardation obtained from measurements made with both dc fields and at 1 kc on Sample B. The dc data is consistently 10% lower than the ac data, indicating that the space-charge effect mentioned earlier is present. The dc measurements were made
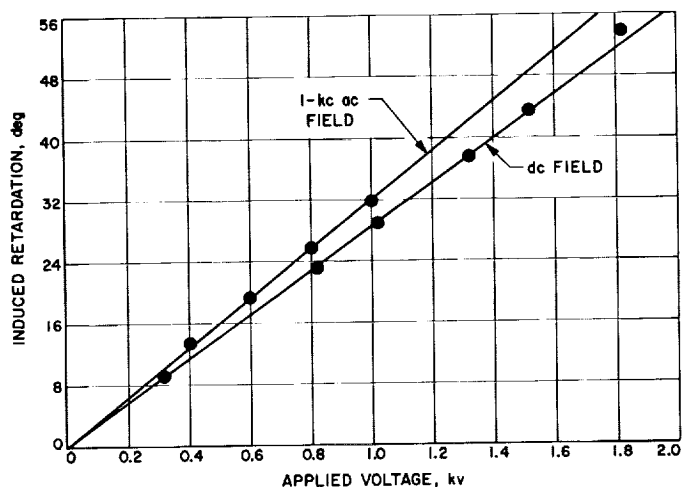
Fig. 15. Linearity of r'_{33} effect, showing space-charge effect seen with applied field

between two consistent sets of ac measurements. Although Meyerhoffer's data (Ref. 20) indicates that the electro-optic effect for c-directed fields becomes very nonlinear within a few degrees of the Curie point, these measurements made with fields to 2000 v/cm indicate the linearity of the effect at room temperature.

Measurements of the induced retardation as a function of temperature were made at 1 kc, with a dc bias of 500 v ($\simeq 1500$ v/cm) to prevent depoling and a 300-v peak ($\simeq 900$ v/cm) ac signal applied across the sample. Using Shumate's data (Ref. 30) for n_a and Meyerhoffer's data (Ref. 20) for the temperature dependence of the spontaneous birefringence δn_s (T), r'_{33} was then computed. The induced birefringence Δn is obtained from the analyzer setting ϕ (Fig. 14) by:

$$\frac{2\pi\Delta n t}{\lambda} = \delta = 2\Delta\phi, \qquad (14)$$

where t is the thickness of the sample along the light path. Then, from Eq. (10),

$$r'_{33} = \frac{2\lambda\Delta\phi}{\pi n_c^3 t \Delta E_c}.$$ (15)

The experimental uncertainty due to that in measuring retardation is about $\pm 3\%$ at room temperature, decreasing to $\pm 2\%$ at 80°C. An additional uncertainty of $\pm 2\%$ in the value of r'_{33} exists due to uncertainty in measuring E_c. Therefore, the total experimental uncertainty is less than 4%. The experimental values of $r'_{33}(T)$ for Samples A and B are shown in Fig. 16. The low-frequency coefficient r'_{33} is 300×10^{-8} cm/statv at room temperature. This is one order-of-magnitude larger than the similar effect in KDP, r_{63}. The effect in barium titanate is compared to that in several other crystals in Table 1.

Table 1. Comparison of the linear electro-optic effect in barium titanate with that in other crystals

Crystal	Crystal class	Low-frequency electro-optic coefficient, cm/statv × 10⁻⁸	Half-wave voltage[a], kv
ZnS[c]	Cubic, $\bar{4}3m$	$r_{41} = 5.9$	10.25[f]
CuCl[d]	Cubic, $\bar{4}3m$	$r_{41} = 18.4$	6.19[f, g]
N₄(CH₂)₆[b, e]	Cubic, $\bar{4}3m$	$r_{41} = 21.9$	9.15[f, h]
KH₂PO₄[d]	Tetragonal, $\bar{4}2m$	$r_{63} = -32, r_{41} = 26$	7.5
KD₂PO₄[d]	Tetragonal, $\bar{4}2m$	$r_{63} = -70$	3.4
NH₄H₂PO₄[d]	Tetragonal, $\bar{4}2m$	$r_{63} = -25, r_{41} = 62$	9.6
BaTiO₂	Tetragonal, $4mm$	$r_{33} \simeq 300$ (20°C) to 1,000 (110°C), $r_{42} \simeq 2250$ (20°C) to 750 (110°C)	0.4 (20°C) to 0.12 (110°C)[f]

[a] $\lambda = 5461$ A.
[b] Hexamethylenetetramine.
[c] Ref. 32.
[d] Ref. 33.
[e] Ref. 34.
[f] Computed for a cubic sample; for Class $\bar{4}3m$, E is along [110]; light propagation is along [110].
[g] Since n_{5461} was not available, n_D was used in the computation.
[h] $\lambda = 5475$ A.

The temperature dependence of r'_{33} can be compared to that of the corresponding dielectric constant K_c. In Fig. 16, K_c, measured on samples obtained from the same source as Samples A and B, has been multiplied by an appropriate constant in order to match it to r'_{33} at 25°C. There is an approximate agreement in form between r'_{33} and K_c, similar to that found in KDP (Ref. 31). Meyerhoffer's work (Ref. 20) indicates that the spontaneous birefringence $(\delta n)_s$ is proportional to the square of the spontaneous polarization. If the induced birefring-

ence also depends on the total polarization, then a field-induced change in $(\delta n)_s$ is given by:

$$\frac{\partial(\delta n)_s}{\partial E} = AP_s\frac{\partial P_s}{\partial E} = AP_sK_c\varepsilon_0,$$ (16)

where A is a constant. Since the induced birefringence is proportional to r'_{33} for a specified field, we might expect that

$$r'_{33}(T) \sim P_s(T)K_c(T).$$ (17)

In Fig. 16, $P_s(T)K_c(T)$ is compared with r'_{33}. It can be seen that the agreement is somewhat improved, but still is not good. There is no immediate explanation for this result.

c. Applied field in the a-direction. Measurements were made on the Linz-type crystal, Sample C. The experimental configuration is shown in Fig. 17. A null occurs when the optic axes of the crystal are aligned with the crossed polarizer and analyzer axes. With a field applied in the a-direction, the oven assembly (and crystal) is
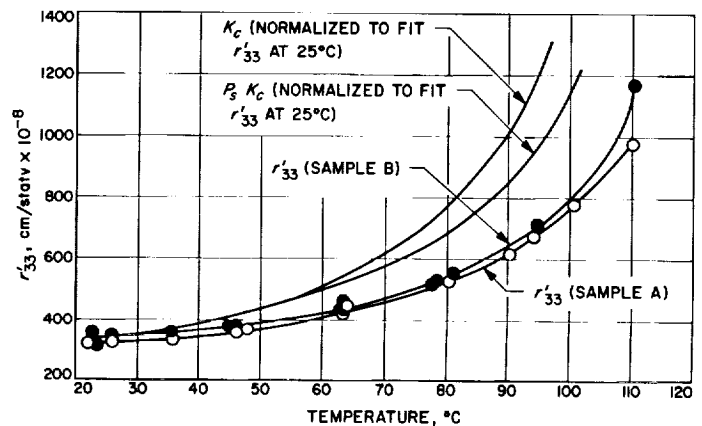
Fig. 16. Plot of r'_{33}, K_c, and P_sK_c vs temperature
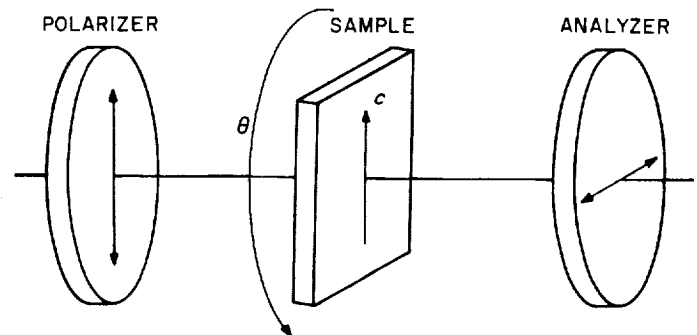
Fig. 17. Experimental setup for measurement of r_{42}

rotated through an angle θ to re-establish the null condition. For fields of $\simeq 1200$ v/cm, θ was of the order of 10 to 30 min of arc, justifying the approximation used in the derivation of Eq. (13). Linearity measurements were made only at room temperature. Fig. 18 indicates that the effect for a-directed fields is also quite linear at room temperature over a large range of applied fields. The values of $r_{42}(T)$ calculated by means of Eq. (13) for Sample C from both ac and dc measurements are indicated in Fig. 19. The ac and dc measurements agree, but there is a 20% uncertainty in the data arising from instrumental noise. At 23°C some discrepancy in the data exists. The larger values were recorded first, indicating some depoling of the crystal may have occurred before the rest of the points were measured.

As with $r'_{33}(T)$ and $\varepsilon_c(T)$, r_{42} exhibits the same general behavior as $K_a(T)$, although the agreement is not quanti-

tative. Again, the electro-optic coefficient is enormous, varying from 2000×10^{-8} cm/statv at room temperature to 750×10^{-8} cm/statv above room temperature. It is interesting to note that the rotation of the total polarization vector, as calculated from a measurement of K_a at room temperature for Sample C, is about 50 min of arc for an applied field of 550 v/cm, corresponding to a rotation of the principal axes of 18 min of arc at the same value of applied field.
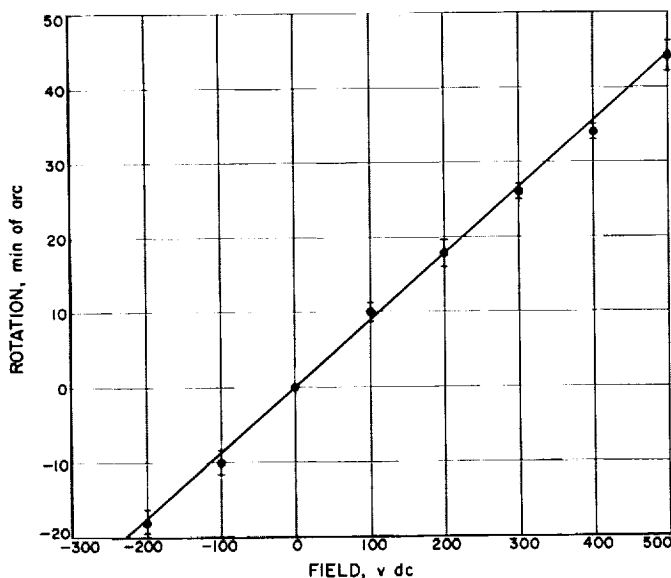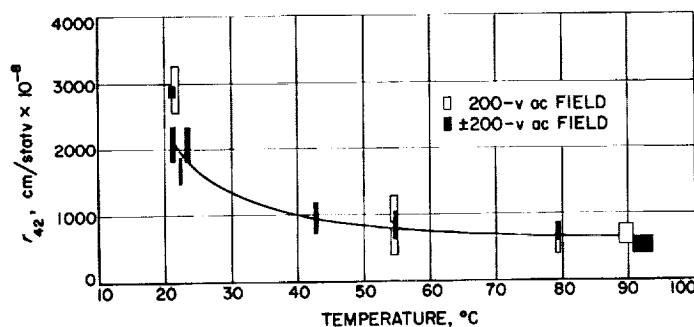
C. Cryogenics Research

J. T. Harding

1. Cryogenic Gyroscope

a. Ac losses in superconducting niobium rotors. Investigation of low-frequency magnetic field losses in superconductors has been under way since their discovery in late 1961 (Refs. 35–37). During the past year three niobium spheres have undergone a series of fabrication and heat treatments in an effort to determine how to eliminate ac losses in cryogenic gyro rotors. An outline of the treatments and the test results are given here.

The spheres were obtained from the Minneapolis-Honeywell Military Products Research Group. The fabrication and heat treatments were done at Minneapolis-Honeywell, while the ac loss measurements were conducted at JPL. Fabrication of the rotors consisted of the following:

(1) Rough machining of an annealed niobium rod (99.9+% pure) to a 1-in. D sphere.

(2) Outgassing of the sphere at 2000°C and 10^{-8} torr.

(3) Lapping and polishing to 5-μin. sphericity.

(4) Vacuum annealing.

Loss measurements after Step (1) were indicative of the ac losses of the niobium prior to any special treatment or processing required for the cryogenic gyroscope application. Successive loss measurements determined the effect of each subsequent step.

Testing for ac losses consists of observing the deceleration rate of the rotor while levitated in a dc magnetic field and spinning at 100 rps. A polar plot showing the magnetic field distribution versus latitude is given in

Fig. 18. Linearity of r_{42} effect

Fig. 19. Plot of r_{42} vs temperature

Fig. 20. Magnetic field on surface of sphere

Table 2. Ac losses at 4.2°K in three niobium spheres

Rotor	Condition	Speed, rps	Deceleration due to ac losses, rps² ±10%	Dissipation[a], μw	Torque[b], dyne-cm
MH2	Rough machined from ingot	——	Calorimetric data, not reproducible		
	Outgassed	——	Calorimetric data, not reproducible		
	Lapped to 5 μin.	100	4.5×10^{-3}	85	1.3
	Annealed	99	6.0×10^{-5}	1.1	1.8×10^{-2}
MH3	Rough machined from ingot	——	Calorimetric data, not reproducible		
	Outgassed	——	Calorimetric data, not reproducible		
	Lapped to 5 μin.	85	1.0×10^{-2}	160	3.0
	Annealed	105	0.90×10^{-4}	1.8	2.7×10^{-2}
MH4	Rough machined from outgassed ingot and annealed	100	4.0×10^{-5}	0.75	1.2×10^{-2}
	Lapped and re-annealed	103	1.3×10^{-5}	0.25	4.0×10^{-3}

[a] $P = I\omega\dot{\omega}$.
[b] $T = I\dot{\omega}$.

Test results are given in Table 2; these data were taken near 100 rps. Data taken over a wide range of rotor speeds indicate that the deceleration rate or the loss of energy per cycle is independent of speed. This implies that the losses are of the hysteresis type and that the rate of dissipation is proportional to frequency. No information regarding dependence on field strength is obtainable since all tests were run in the same field.

b. Gyro tests. The discovery of ac losses ended all efforts to test the feasibility of the cryogenic gyroscope. Recently, as discussed above, ac losses have been reduced to the point where spindown times are extended to days or months. This makes possible, for the first time, the observation of a spinning rotor for a sufficient time to demonstrate behavior as a gyroscope, using the Earth as a rate table. Drift of the rotor spin axis relative to inertial space has been measured for each of the outgassed rotors discussed above. Results are very encouraging in that drift rates substantially below the Earth's rate have been observed.

The system used to support and house the rotor is shown in Fig. 21. The glass housing is double-walled. Helium gas for spinup is conducted through the annular region past the radiation and cooling baffles and exits through jets located around the levitated rotor. After operating speed is reached, a high vacuum can be produced within the housing by pumping on the inner chamber. Two niobium wire-wound coils carrying current in

Fig. 20. The symmetry axis of the levitational field is vertical, while the rotor spin axis lies in the horizontal plane. Formerly, ac losses were measured calorimetrically, i.e., by observing the time required for a thermally isolated superconductor to heat up to its transition temperature in a uniform ac magnetic field. At low loss rates, thermal isolation was found to be inadequate and desorption was a problem. The deceleration technique, wherein energy dissipation is calculated as the decrease in kinetic energy, is free of complicated side effects and has yielded reproducibility for loss rates of 2.5×10^{-7} w.

In order to keep the specimen temperature constant while absorbing energy, it is necessary to admit some helium gas into the space surrounding the rotor so that heat can be conducted to the liquid helium bath. At the same time, the gas produces drag. The deceleration caused by the gas can be made a small fraction of the total deceleration and yet provide adequate cooling for rotor speeds less than 200 rps. However, a correction must be made; this is done by measuring the deceleration at 10^{-4} torr and making use of the linear dependence of gas drag on pressure in the Knudsen range.

Fig. 21. Cryogenic gyro

Fig. 22. Cryogenic gyro in Dewar

opposing directions create the gradient field required to support the rotor. The leads to these coils are shorted so that they can carry a persistent current. No electric power is required once the rotor is levitated. The assembly is contained in a stainless-steel double Dewar which contains liquid helium for maintaining the gyro at 4.2°K (Fig. 22).

Operation consists of spinning the rotor about a horizontal axis in an approximately east-west direction. During the course of a day as the Earth rotates, the axis of the ball appears to rotate with respect to the Earth-fixed Dewar. Neglecting drift, the rotor spin axis will trace a full circle with respect to the Earth about the

Earth's polar axis during one sidereal day. The rotor axis is viewed through a port in the Dewar with a microscope attached to the Dewar. Due to microscopic irregularities of the rotor surface, the poles are visible as stationary points at the center of concentric circles. Since the rotor is a solid sphere, this method of readout eliminates any need for a preferred spin axis. Because the microscope has limited freedom of motion, determination of drift can only be made when the axis is within the viewing port. In practice the microscope crosshairs are set on the axis at an initial time, and subsequent readings are made at approximately 12-hr intervals whenever a pole crosses the crosshairs. The crossing times can be measured with an accuracy of ±1 min in time, which represents an uncertainty of ±15 min of arc.

In order to interpret the drift data presented in Table 3, it is necessary to understand the nature of the observed drift. The trajectory of the spin axis in inertial space is very nearly a cone about the Earth's polar axis. Since the spin axis is initially almost perpendicular to the

Table 3. Drift data for three rotors

Rotor	Speed range, rps	Length of test, hr	Average absolute drift rate, deg/hr	Average torque required to produce drift[a], dyne-cm	Static period, sec
MH2	251–219	71.5	0.056	0.020	25
MH3	240–196	92.3	0.562	0.19	45
MH4	99–88	49.5	3.13	0.45	20

[a] $\tau = I\omega\Omega$.

Earth's polar axis in the present test, the observed drift is an east- or west-traced circle. Strictly speaking, the circle is not perfect; actually the trajectory has a slight spiral toward Polaris. Drift data for the three rotors are presented graphically in Fig. 23.

The period of oscillation of each rotor was measured when first levitated by rotating it slightly from its equilibrium orientation and observing the ensuing motion. The period was again measured after the gyro test when the rotation had stopped. In each case there was no significant change in period as a result of spinup and spindown. The static period provides an upper limit on drift rate by the following relationship:

$$\left(\frac{2\pi}{T}\right)^2 > \omega\Omega,$$

where T is the static period, ω is the rotor speed, and Ω is the precession rate. The drift rate is considerably below the limit because components of torque perpendicular to the spin axis average out due to spin.

No attempt was made in these tests, neither mechanically nor analytically, to compensate the errors contributing to drift. Although the rotors were lapped to high sphericity and then annealed, they were not specifically mass-balanced. The magnetic field was not adjusted to reduce drift; the field was predetermined and remained the same for all tests. The results are presented as raw data, and none have been omitted.
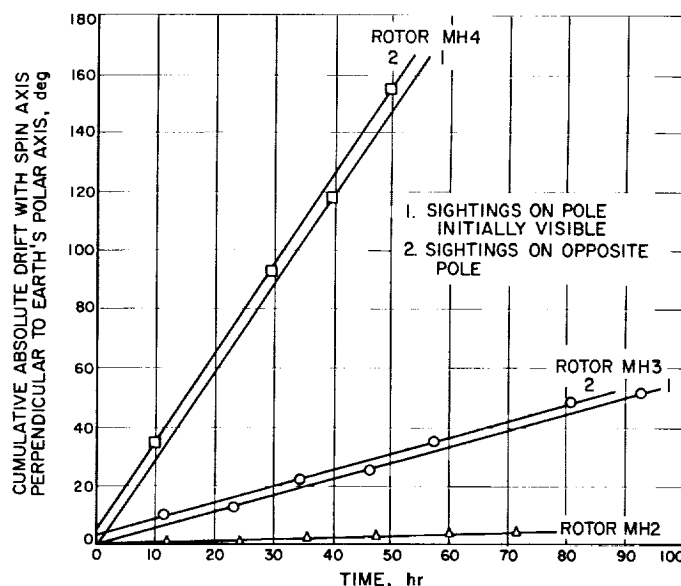
Fig. 23. Drift vs time for three niobium rotors

It is felt that these results demonstrate the feasibility of the cryogenic gyro concept. Not only are absolute drift rates well below the Earth's rate attainable, but, to a large extent, the drift is predictable. The major causes of drift are the same factors that give rise to a finite static period: non-sphericity and trapped flux. (In the case of Rotor MH2, the London moment induced in the spinning superconductor is an important source of drift, about 1.1 deg/day.) The sphericity of the rotors can easily be improved by a factor of 2 to 3. Their asphericity is over 5 μin., which is the value obtained after lapping and before final annealing. The sphericity of the rotors will be measured after the present gyro tests. Trapped flux can easily be reduced by 2 orders-of-magnitude by the use of additional μ-metal shielding during cooldown. In the research model gyro, there are a number of design features which have not been optimized. In a development model, the rotor size would be at least twice as large. If possible, the rotor would be fabricated as a hollow shell, rather than the solid ball used here. The rotor speed should be increased to as high a value as the rotor will tolerate, i.e., 1000 rps or more. All of these factors could reduce drift by approximately 2 orders-of-magnitude.

References

1. Smit, J., Thesis, University of Leiden, Belgium, April 1956.

2. Parker, R., "An Analysis of the Magnetization Processes in Iron Single Crystals by an Electrical Method," Philosophical Magazine, Vol. 1, p. 1133, 1956.

3. Kuwahara, K., "Magnetoresistance Effect in Aluminum-Iron," Journal of Science of Hiroshima University, Vol. A22, p. 267, 1958.

4. Goto, M., Hiraoka, T., and Kuwahara, K., Journal of Science of Hiroshima University, Vol. A-II 26, p. 117, 1963.

5. Humphrey, F. B., "Very Thin Permalloy Films," SPS 37-18, Vol. IV, pp. 64–66, Jet Propulsion Laboratory, Pasadena, California, December 31, 1962.

6. Seavey, M. H., and Tannenwald, P. E., "Ferromagnetic Resonance in Ultra Thin Films," Journal of Applied Physics, Vol. 29, p. 292, 1958.

7. Nishimura, A., Hiraoka, T., and Kuwahara, K., "Spontaneous Magnetization and Hall Effect in Thin Films of Ni-Fe," Journal of Science of Hiroshima University, Vol. A-II 26, p. 113, 1963.

8. Humphrey, F. B., "Magnetic Character of Very Thin Permalloy Films," Journal of Applied Physics, Vol. 34, p. 1067, 1963.

9. Verderber, R. R., "X-Ray Fluorescence Analysis of the Composition of Ni-Fe Films," Norelco Reporter, Vol. 10, p. 30, January 1963.

10. Humphrey, F. B., and Johnston, A. R., Sensitive Automatic Torque Balance for Thin Magnetic Films, Technical Report No. 32-351, Jet Propulsion Laboratory, Pasadena, California; reprinted in Review of Scientific Instruments, Vol. 34, No. 4, pp. 348–358, April 1963.

11. Humphrey, F. B., Johnston, A. R., and Emmett, J. L., "Torsion Magnetometer," RS 36-11, pp. 16–18, Jet Propulsion Laboratory, Pasadena, California, November 1, 1961.

12. Bozorth, R. M., Ferromagnetism, D. Van Nostrand Company, Inc., New York, 1961.

13. Takahashi, M., et al., "Induced Magnetic Anisotropy of Evaporated Ni-Fe Films," Journal of the Physical Society of Japan, Vol. 16, p. 1351, 1960.

14. Néel, L., "Anisotropic Magnétique Superficielle et Surstructures d'Orientation," Journal of the Physics of Radium, Vol. 15, p. 225, 1954.

15. Taniguchi, S., "A Theory of the Uniaxial Ferromagnetic Anisotropy Induced by Magnetic Annealing in Cubic Solid Solutions," The Science Reports of the Research Institutes, Tohoku University, Series A, No. 7, p. 269, 1955.

16. Marechal, M. J., "Contributions à l'Etude Experimentale des Surstructures d'Orientation Magnétiques des Alliages Fer-Cobalt," Journal of the Physics of Radium, Vol. 16, p. 122S, 1955.

References (Cont'd)

17. Chikazumi, S., and Oomura, T., "On the Origin of Magnetic Anisotropy Induced by Magnetic Annealing," *Journal of the Physical Society of Japan*, Vol. 10, p. 842, 1955.

18. Heindenreich, R. D., and Reynolds, F. W., "Uniaxial Magnetic Anisotropy and Microstructure of Ferromagnetic Metal Films," *Structure and Properties of Thin Film*, John Wiley & Sons, Inc., New York, p. 402, 1959.

19. Hansen, M., *Constitution of Binary Alloys*, McGraw-Hill Book Company, Inc., New York, p. 472, 1958.

20. Meyerhoffer, D., "Transition to the Ferroelectric State in Barium Titanate," *Physical Review*, Vol. 112, p. 413, 1958.

21. Billings, B. H., "The Electro-optic Effect in Uniaxial Crystals of the Type XH_2PO_4: Part I, Theoretical," *Journal of the Optical Society of America*, Vol. 39, p. 802, 1949.

22. Nye, J. F., *Physical Properties of Crystals*, Oxford University Press, p. 250, 1957.

23. Sterzer, F., Blattner, D., and Miniter, S., *Journal of the Optical Society of America*, Vol. 54, p. 62, 1964.

24. Weingart, J. M., "Electro-optical Properties of Barium Titanate," *SPS 37-18*, Vol. IV, pp. 68–71, Jet Propulsion Laboratory, Pasadena, California, December 31, 1962.

25. Weingart, J. M., and Johnston, A. R., "An Electronic Polarimeter for Measurement of the Electro-optic Coefficients of Barium Titanate," *SPS 37-22*, Vol. IV, pp. 31–33, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.

26. Remeika, J. P., "A Method for Growing Barium Titanate Single Crystals," *Journal of the American Chemical Society*, Vol. 76, p. 940, 1954.

27. Johnston, A. R., "Poling of $BaTiO_3$ Samples," *SPS 37-20*, Vol. IV, pp. 24–29, Jet Propulsion Laboratory, Pasadena, California, April 30, 1963.

28. Partington, J. R., *An Advanced Treatise on Physical Chemistry*, Vol. IV, Longmans Green and Co., New York, p. 156, 1953.

29. Triebwasser, S., "Space Charge Fields in $BaTiO_3$," *Physical Review*, Vol. 118, p. 100, 1960.

30. Shumate, M. S., and Johnston, A. R., "A Method for Determination of the Refractive Index and Its Application to Flux-Grown Barium Titanate," *SPS 37-22*, Vol. IV, pp. 36–39, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.

31. Zwicker, B., and Scherrer, P., *Helvetica Physica Acta*, Vol. 17, p. 346, 1944.

32. Namba, S., "Electro-optical Effect of Zincblende," *Journal of the Optical Society of America*, Vol. 51, p. 76, 1961.

33. *American Institute of Physics Handbook*, 2nd Edition, distributed by McGraw-Hill Book Company, Inc., New York, 1963.

References (Cont'd)

34. McQuaid, R. W., "The Pockels Effect of Hexamethylenetetramine," *Applied Optics,* Vol. 2, p. 320, 1963.

35. Harding, J. T., and Tuffias, R. H., "Measurement of AC Losses in Superconductors," *RS 36-14,* pp. 22–26, Jet Propulsion Laboratory, Pasadena, California, May 1, 1962.

36. Harding, J. T., and Tuffias, R. H., "Alternating Magnetic Field Losses in Superconductors," *SPS 37-16,* Vol. IV, pp. 105–108, Jet Propulsion Laboratory, Pasadena, California, August 31, 1962.

37. Harding, J. T., and Tuffias, R. H., *AC Losses in Superconductors,* JPL Section 345 Technical Report No. 345-4, Jet Propulsion Laboratory, Pasadena, California, October 1962.

ENGINEERING MECHANICS DIVISION

V. Materials Research

A. Pure Oxide Ceramic Research

M. H. Leipold

The mechanical behavior of polycrystalline oxide materials is markedly affected by the presence of minor amounts of impurities (Refs. 1, 2, 3, 4). At high temperature, deformation is enhanced by the formation of minor amounts of easily deformed material at the grain boundaries resulting in easy slide of one grain past another. The concept of deformation of grain boundaries in ceramic oxides has been supported by work in the literature (Refs. 3 and 4).

It is of interest to determine the inherent mode of deformation of these materials when the effect of impurities has been eliminated. Because of the unavailability of oxide materials with very low impurity content such information is not available.

Other properties of refractory oxides may be affected by very small amounts of impurities. For example, it has been reported that small amounts of impurity may result in an increase in porosity when dense oxides are exposed to very high temperatures (Ref. 5). Such a suggestion could be confirmed by examination of materials not containing these impurities.

In order to make available dense polycrystalline ceramic oxide specimens containing very low levels of impurity it was necessary, first, to develop powders containing as little impurity as possible and, secondly, to develop techniques for fabricating these powders into usable specimens. A possible solution to the first problem, that of production of very pure powder, was obtained[1] and is being presently evaluated. In brief the technique involved solution of high-purity magnesium metal, extraction of impurity by means of a selective chelating agent, and precipitation of magnesium carbonate. This precipitate was then calcined under vacuum to produce very high-purity, fine magnesium oxide powder.

---

[1]Sample from R. A. Weeks, Oak Ridge National Laboratory.

LOAD

WATER-COOLED RAM

SIGHT PORT

REFRACTORY SPACER

RAM

WATER-COOLED FURNACE SHELL

DIE

SPECIMEN

THERMOCOUPLES

RESISTANCE HEATING ELEMENT
(I OF 8 SHOWN)

12 in.
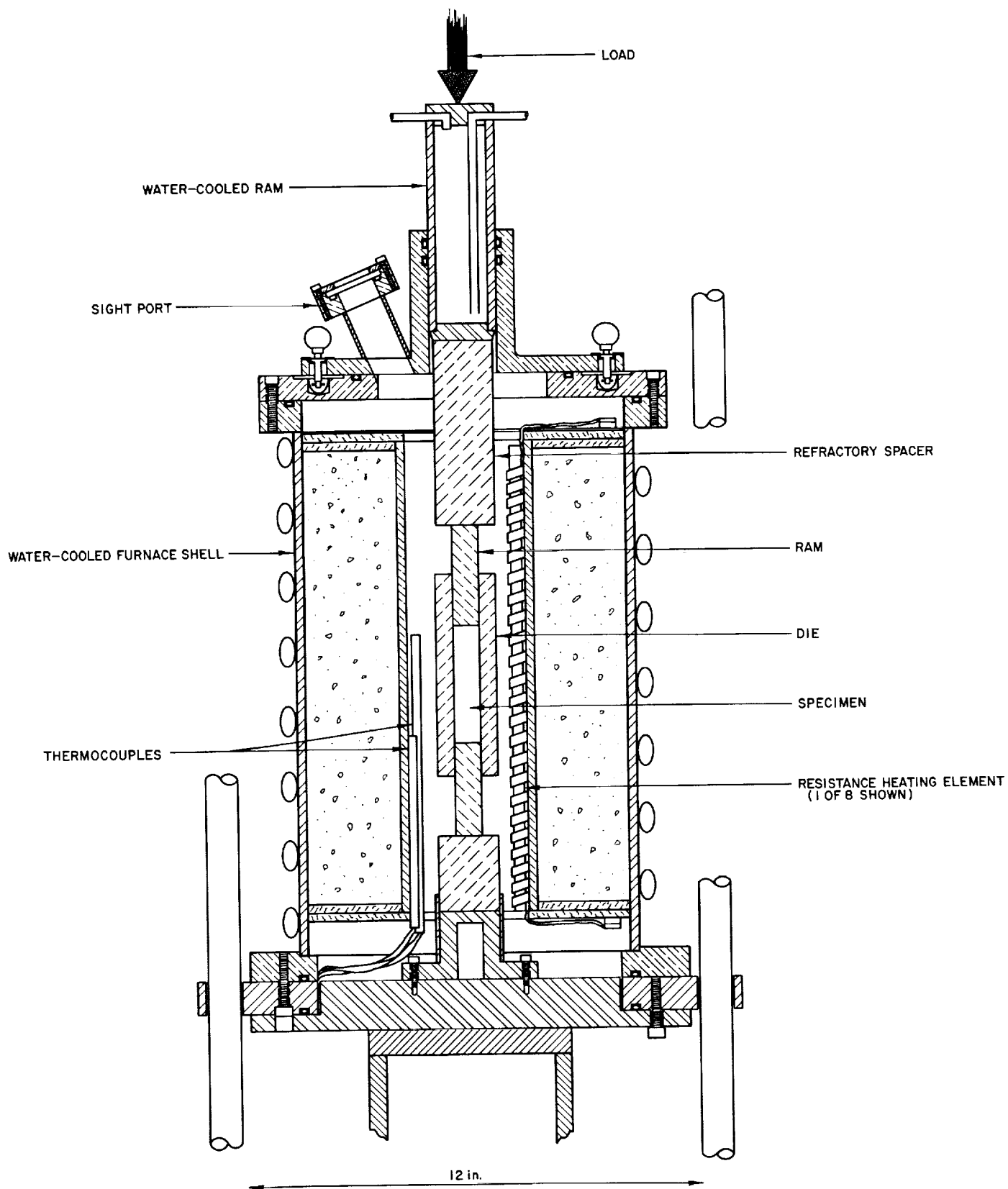
Fig. 1. Cross-section of oxide hot-press chamber

The second step in the production of dense polycrystalline specimens was a technique which will permit densification of the fine pure powders without the introduction of additional impurity. One of the best techniques available was hot pressing. However, one serious problem with the normal graphite die hot pressing has been the introduction of impurity from the graphite dies and, to some extent, introduction of graphite itself to the oxide powders. A modification of this standard hot-pressing procedure which to a large extent will eliminate these problems was the use of aluminum oxide dies and rams at lower temperatures. The higher strength available in the aluminum oxide material permitted the use of higher pressing pressures and consequently produced theoretically dense material at lower pressing temperatures. The combination of greater compatibility between the oxide dies and rams with the oxide powder and lower temperatures greatly reduced the tendency toward contamination. In addition, the use of lower pressing temperatures resulted in a material having an extremely fine grain structure. Oxide die hot-pressing technique has been reported in the literature (Ref. 6).

An oxide die hot-pressing facility has been designed and constructed at JPL and is shown in Fig. 1. The unit may be operated under a vacuum of the order of 50 μ or a variety of gas atmospheres including pure oxygen. Provision was made for improved vacuums if required. The unit has a maximum operating temperature of 1200°C and loads up to a maximum of 40,000 lb may be applied. Linear heating and cooling rate were automatically controlled.

A critical portion of an oxide die hot-pressing system is the choice of die material. Aluminum oxide is the most widely used because of the relatively high strength at 1100°C (approximately 30,000 psi). The most successful dies used to date have been those which have been hot pressed to 99% theoretical density at this laboratory from Linde A5175 alpha alumina (Ref. 5). Some attempts have been made to use a high-grade commercial aluminum oxide. This material had comparable purity to the JPL material; however, the density was about 93% of theoretical. The commercial aluminum oxide was not nearly so successful a die material, failing every time under normal use conditions. Further, entirely different modes of failure were encountered between the two materials. The commercial aluminum oxide failed quite docilely into two pieces, while the JPL pressed material shattered when it did fracture, scattering many pieces throughout the furnace.

Attempts have been made to support the commercial aluminum oxide dies by means of ⅛-in. thick by 1-in. wide molybdenum rings (Climax Corporation, Grade TZM) which are ground to fit the outer surface of the die case. The size of the ring is adjusted so that at 1100°C all of the load will be theoretically supported by the molybdenum ring. This technique has been successful and has reduced die failures using the commercial material to approximately one in ten. However, use of the die is limited to a vacuum or inert atmosphere.

Some experimentation is also under way using other die materials. Molybdenum TZM dies have been successfully used; however, galling occurred between the die and rams and the components could not be separated without a fracture. Modifications of the technique are under consideration. Also refractory carbide dies may be used, both for the oxides and for use in the high purity carbide research program.

Specimens produced by the oxide die hot pressing have been limited to small pieces of material, while the technique was developed and knowledge of pressing parameters determined. Figs. 2 and 3 show micro and macro
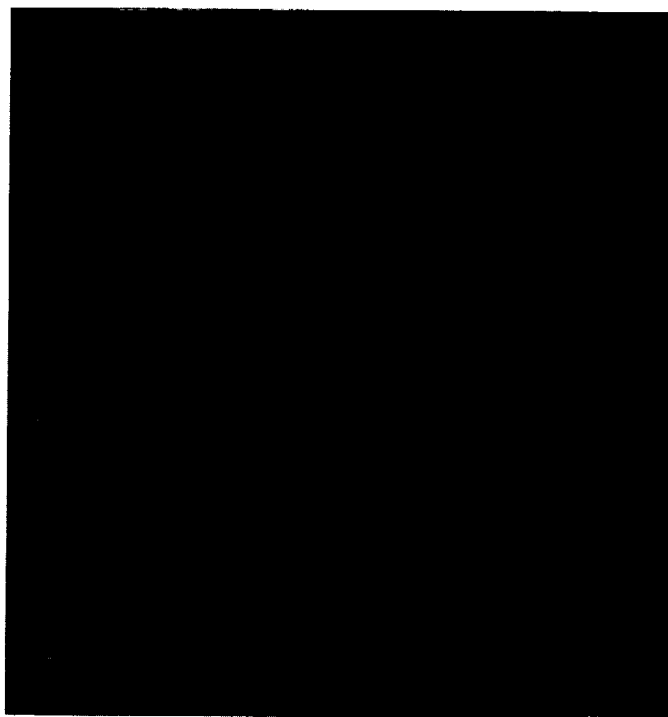
Fig. 2. Electron micrograph of oxide die hot-pressed MgO. Black lines are 1 μ apart.
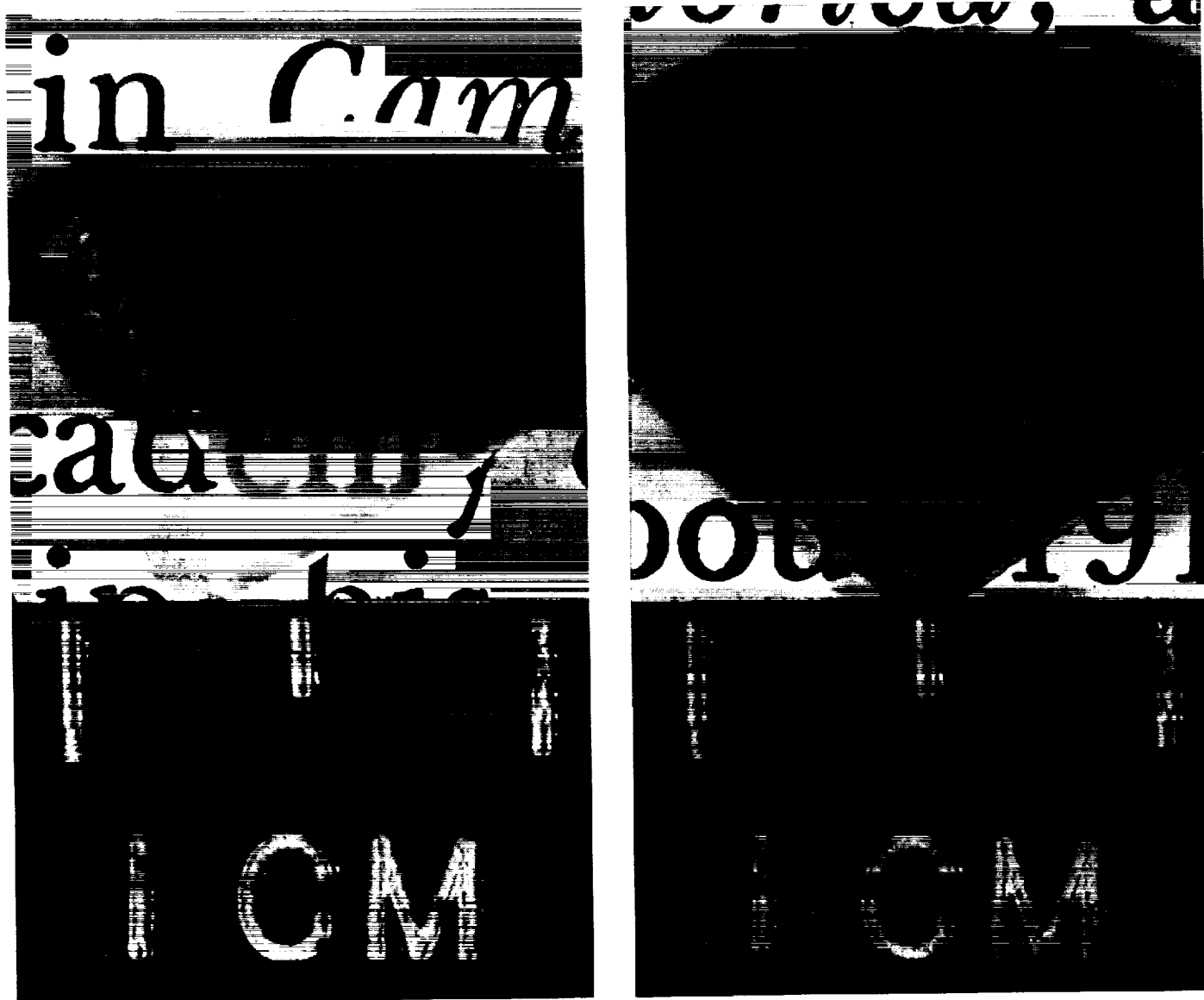
Fig. 3. Oxide die hot-pressed MgO, 2 mm thick. Specimens polished on both sides.

views of oxide die hot-pressed magnesium oxide produced from commercial 0.1-μ starting material. The electron micrograph (Fig. 2) was prepared by Sloan Research Industries, using the Parlodion replica technique. The specimen was polished and steam etched. Both specimens were pressed at 1150°C, 15,000 psi for one-half hour. The left specimen (Fig. 3), which is colorless, was pressed under vacuum, while the right specimen, which is pink in color, was pressed in air. The difference in color has been at present attributed to the oxidation state of minor impurity within the materials, and is presently under further investigation. The starting material of each of these specimens, Fisher Grade M300 magnesium oxide, was

reported approximately 99.5% pure, containing 0.1 to 0.2% each of calcium, silicon, and iron as the principal known impurities.

Several pressings have been made using the higher purity magnesium oxide produced at this laboratory. These have generally been successful and have indicated that pressings may be accomplished at lower temperatures than required with the M300 material. This would be a result of the finer particle size obtained when the high purity MgO is calcined in vacuum at temperatures in the 400 to 800°C range. Early hot pressing tests made

with a sample of high purity MgO calcined 900°C in air[2] were not successful. It was necessary to substitute vacuum calcining for the previously used air calcining to obtain powders which were sufficiently fine to permit fabrication into dense specimens in the oxide die hot press. The material calcined in air was composed of particles 1 to 2 μ D, while the vacuum calcined material should yield particles in the range of 0.01 μ. Some problems have been encountered with incomplete calcination in the high purity magnesium oxide. These problems have taken the form of delamination of the pressed specimens, and failure to attain theoretical density. Investigation of this material will continue.

---

[2]R. A. Weeks, Oak Ridge National Laboratory, private communication.

B. Pure Carbide Ceramic Research

M. H. Leipold

The work reported in Ref. 7 on the production of dense pure tantalum carbide specimens is continuing with most of the effort being concentrated on the purification of available raw materials. Heating tantalum containing 0.1 wt % oxygen in a high vacuum lowered the oxygen content; however, the finely divided tantalum powder sintered into a solid and could not be used in the production of tantalum carbide without particle size reductions. Particle size reduction of a malleable metal such as tantalum was felt to be difficult without introduction of additional impurities.

A second purification approach was attempted by dissolving hydrogen into the tantalum metal powder at 450°C and 0.1 torr hydrogen pressure. It was hoped that subsequent removal of the hydrogen would remove the oxygen impurity. The hydrogen used had been purified by diffusing it through palladium at 325°C. The hydrogen was removed from the tantalum at 800°C in a vacuum of 2 × 10⁻⁶ torr. Chemical analysis indicated no reduction in oxygen levels as a result of this treatment.

A third technique employed carbon to remove oxygen from the tantalum and was successful. Thus the carburi-

zation of the tantalum to form tantalum carbide powder and the removal of oxygen could be accomplished simultaneously. By this technique tantalum carbide powder has been produced with oxygen levels of approximately 0.03 wt %. This carbide powder has been produced from National Research Corp. Grade SGQ tantalum powder and from Speer Carbon Co. No. 11 nuclear flour. The graphite flour had been degassed at 1200°C in vacuum of 2 × 10⁻⁶ torr before use and stored in a dry helium atmosphere. These materials were reacted in vacuum of the order of 5 × 10⁻⁶ torr at temperatures above 1400°C. After 8 hr at 1800°C, a stoichiometric mixture of carbon and tantalum had proceeded to 92% completion.

It had been noted during the investigation that merely exposing the finely divided carbide powder to the air results in additional oxygen and nitrogen contamination. In one case the oxygen level changed from 0.03 to 0.09 wt % and in another case from 0.09 to 0.3 wt % after 48 hr in air. Nitrogen increases were also evident although not so pronounced. These differences in sensitivity to air contamination may be a result of uncontrolled differences in particle size of the carbide powder.

These purification techniques have resulted in the capability of producing tantalum carbide powder containing approximately 0.03% oxygen, 0.015% nitrogen, and total metallic impurities of approximately 0.01%. Such materials should be satisfactory for initial investigation of mechanical properties.

During the investigations reported here a commercial source of tantalum carbide became available which may be capable of approaching, although at present not equalling, the levels of purity attained here. This source will be evaluated and the possibility of obtaining material purified to a still greater extent will be investigated.

The second phase in obtaining usable tantalum carbide specimens for evaluation is the hot pressing of this high purity powder into dense polycrystalline blanks. A vacuum hot press has been purchased and is presently being installed. This unit is designed for maximum operating temperatures of 2500°C under inert gas or a vacuum. The vacuum would be of the order of 10⁻⁵ torr below 2100°C. Above this temperature the volatility of carbon limits the vacuum attainable. With this hot press it is expected that specimens closely approaching theoretical density can be fabricated without introduction of additional contamination.

C. Graphite

W. V. Kotlensky

1. Glassy Carbon

Preliminary tensile results on glassy carbon were reported previously (Ref. 8). Additional tests have been made for the purpose of delineating its high temperature tensile behavior.

The glassy carbon was supplied by the Tokai Electrode Manufacturing Company. Two different lots were tested. These were reported by the manufacturer to have been heat-treated at 2000°C (Lot 228) and 3000°C (Lot 229) and to have a density of 1.50 g/cm³. Tensile results are given in Table 1.

As seen in this table the strengths of these two lots are not markedly different. The scatter in the data is not unexpected for such a new material. Over the temperature range studied, both lots are from 2 to 4 times stronger than standard pitch-coke graphites, and show the same behavior as pitch-coke graphites of increasing strength up to 2500°C, followed by a dropoff in strength above this temperature. When compared on a strength-to-density ratio basis, the strength of glassy carbon appears more remarkable. At 2500°C, for example, glassy carbon is as strong as pyrolytic graphite.

The ductility at 2500°C and above as evidenced by the recorded elongation for Lot 228 (the material heat-treated at 2000°C) is similar to pitch coke graphite, and

Table 1. Tensile properties of glassy carbon

Test Temperature, °C	Lot 228, heat-treated at 2000°C		Lot 229, heat-treated at 3000°C	
	Ultimate strength, psi	Recorded elongation,[a] %	Ultimate strength, psi	Recorded elongation,[a] %
Room temp.	5,400	—	5,900	—
1600	8,000	0.5	8,100	0.7
1600	6,700	0.15	6,400	0.6
1900	11,300	1.2	11,800	1.1
1900	16,100	2.5	10,000	1.3
2200	17,500	2.3	12,100	1.1
2200	16,700	2.6	14,800	3.6
2500	20,800	13.2	14,400	2.1
2500	20,100	9.0	20,800	3.4
2500	—	—	25,300	3.8
2700	16,100	23.2	20,400	5.4
2700	17,700	33.1	16,400	1.8
2900	10,400	12.2	14,200	2.5
2900	12,700	23.5	15,500	3.9

[a]Corrected for deformation in the filleted regions.
Strain rate ~ 2 X 10⁻⁴ in./in./sec.

is markedly greater than Lot 229 (the material heat-treated at 3000°C). The lack of appreciable ductility in the Lot 229 glassy carbon can be attributed to the 3000°C heat treatment, since this is the only reported difference between this lot and Lot 228. Continuing studies are being made on the effect of heat treatment of Lot 228, tested at temperatures above 2000°C, as well as on structural changes of the heat-treated and deformed materials.

References

1. Gilman, J. J., "Dislocation Sources in Crystals," *Journal of Applied Physics,* Vol. 30, 1959, p. 1584.

2. Stokes, R. J., "Dislocation Sources and the Strength of Magnesium Oxide Single Crystals," *Transactions of the Metallurgical Society of AIME,* Vol. 224, 1962, p. 122.

3. Stokes, R. J., "Microstructure and Mechanical Properties of Ceramics," Minneapolis-Honeywell Co., Report 63-264, 1963, p. 42.

References (Cont'd)

4. Folweiler, Robert C., "Creep Behavior of Pore-Free Polycrystalline Aluminum Oxide," *Journal of Applied Physics,* Vol. 32, 1961, p. 773.

5. Nielsen, T. H., and Leipold, M. H., "Thermal Expansion in Air of Ceramic Oxides to 2200°C," *Journal of the American Ceramic Society,* Vol. 46, 1963, p. 381.

6. Spriggs, R. M., et al., "Hot Pressing Ceramics in Aluminum Dies," *Bulletin of the American Ceramic Society,* Vol. 42, 1963.

7. Leipold, M. H., and Nielsen, T. H., "Ceramics," *Space Programs Summary No. 37-23,* Vol. IV, October 31, 1963, pp. 51–52.

8. Kotlensky, W. V., "Graphite," *Space Programs Summary No. 37-24,* Vol. IV, December 31, 1963, p. 46.

VI. Electromechanical Engineering Support

A. The Packaging of Integrated Circuits

L. Katzin

The use of integrated circuits (chips[1]) has the following advantages: reduction in volume, weight, and power consumption; improved reliability; increase in frequency; and reduction of design and fabrication time.

1. Definition of Problem

As low-level chip assemblies are combined to form larger portions of a system, the packaging density decreases because of large unusable voids in the resulting

geometry. In four integrated circuit modules being built for JPL, the volumetric efficiency varies from 2 to 5%.

Chips have a higher inherent reliability than corresponding circuits made from discrete conventional components. The assembly of chips requires a higher level of human dexterity than the assembly of conventional components. Because of the density of interconnections, packaging engineers attempt to increase interconnection area by designing in three dimensions. With the increased complexity, the reliability of the assembly may not offer as great an advantage as the reliability of the chip indicates.

An additional problem which arises with tight multi-layer packaging of chips is the limited capability for change and repair and the resulting reliability degradation when repairs or changes are made.

The packaging problem is being studied with the objectives to:

(1) Decrease the human element in assembly.

(2) Allow for repair and change with reasonable accessibility.

(3) Reduce design and fabrication time.

---

[1] A chip is a miniature circuit formed of silicon, typically 0.005 in. thick and 1/16 in. long. Sunk into the chip in gaseous form, under high temperatures, are microscopic particles of impurities (such as boron or phosphorous) which act as transistors, diodes, and other electronic components. The tiny particles are connected by a thin coat of metal (usually aluminum) which is vaporized and condensed onto the surface of the silicon, where it is etched into a circuit pattern. The final product is encased in a metal or ceramic container, with little metal connections protruding from it.

(4) Reduce number of joints.

(5) Eliminate friction contacts in flight.

Three complementary studies are now in various stages to develop: (1) a technique for interconnecting chips on a modular level, (2) a method for interconnecting modules, and (3) a microminiature connector.

2. Interconnection of Chips

The technique for interconnecting a group of ten to thirty chips into a module is based on a system of laminated preforms.[2] This system permits complex interconnections with a minimum of junction interfaces, a

---

[2]Katzin, L., "The Interconnection of Integrated Circuit Chips," JPL interoffice memo, January 1964.
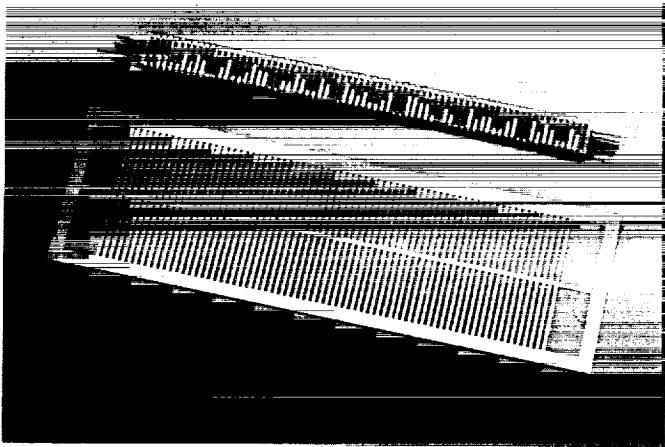
Fig. 1. Models of interconnection sticks

practical minimum of the human element in assembly, and very little lead time. Maintenance, repair and change capability are possible at the component level. The preforms being used in the initial mockups are "fish skeletons" of thin beryllium copper, each of which represents a circuit nodal point. These preforms can be cut at the main trunk as often as necessary, thus establishing more circuit nodes (Figs. 1 and 2). Ribs are removed as needed, and the preforms are laminated one on top of the other, separated by ribbons of insulation. The lamination when finished will resemble an original preform, with only one rib extending from each side at any position and a greater trunk (spine) thickness. The ribs are then folded up and the entire lamination is placed into a plastic trough.

Chips are placed on top of the trough such that their leads are superimposed upon the ribs of the preformed laminate. The leads are then welded to the ribs with a conventional resistance or gap-welding technique. There is only one weld per lead to interconnect all of the chips on the trough (or stick). A stick 0.200 × 0.150 × 4.50 in. is capable of interconnecting fifteen Texas Instruments chips, permits 30 input and output leads, occupies 0.125 in.[3] and has a volumetric efficiency of approximately 10%.

As mentioned, the first prototypes have been limited to a stick geometry; however, it appears quite reasonable to use the same technique in a planar orientation, which would permit more chips in a module assembly (a 3 × 3 in. module could accommodate 120 chips, still requiring only one weld per used chip lead).
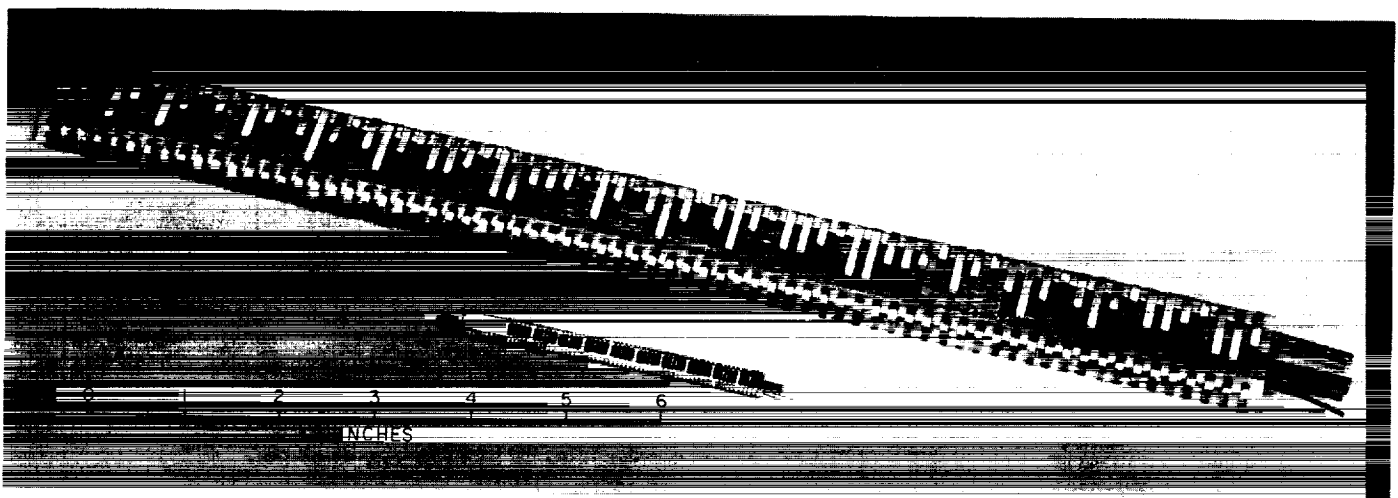
Fig. 2. Interconnection stick and nodal preform

The technique as described requires very little design time, as the circuit information can be used in tabulated form and the removing of ribs and cutting of trunk can be pseudo-automated with graphic arts, or totally automated by a punched tape.

One trial stick has been built, and simple tooling has been completed for a second model. This general technique has been adopted by Fairchild Space and Defense Systems in performing a contract for JPL.

3. Module Interconnection

An effort is under way to develop a technique of welding magnet wire through the insulation. The intended application for this process is a back panel or "mother board" which will permit high density point-to-point wiring from a programmed instruction. This is *not* automation for cost reduction, but rather for control (which means reliability) and time saving (which means flexibility).

The intent is to interconnect a miniature high density terminal board point-to-point on centers of 0.032 to 0.050 in. in so little time that rather than changing the interconnect, a new revised panel can be built to replace the old one. This will provide on a 3 × 2½ in. panel, interconnect capacity for ten 3 × 3 in. modules of 120 chips packaged as described above, or a totally interconnected module of 1,200 integrated chips in a block of 3½ × 3 × 2½ in. (26¼ in.^{3}) or about 4% volumetric efficiency on a system level.

This technique would also simplify and increase the reliability of memory array terminations, where magnet wire is now used. Preliminary investigation and experimentation resulted in very strong confidence in the ability of industry to solve this problem without a significant change in today's state of the art.

4. Microminiature Connector

A microminiature connector is in its second stage of development. The simplicity of design offers potentially very high reliability with a selection of wide range of insertion and withdrawal forces. The principle of this connector differs from that of conventional connectors in that the female is rigid and the male element is deflected.

The male pin is gold-plated beryllium copper spring wire, 0.016-in. D. This pin is engaged into a seamless brass tube 0.020 I.D. with 0.006-in. wall, with a slight "S" bend in the shank. This bend is approximately an 0.020-in. offset, which is sufficient for a good mating action, yet not sufficient to exceed the elastic limit of the male pin. The adjustment in insertion and withdrawal force can be made at the time of manufacture by varying the amount of offset.

This connector could be used with the stick assembly and/or the "mother board" to provide pressure contact assemblies through test and checkout and later converted to a welded joint which could be reconverted to a pressure contact if necessary.

B. Welded Electronic Packaging

R. M. Jorgensen

Recent efforts in welded electronic packaging have been divided between providing JPL with a facility capable of accomplishing flight quality rework and fabrication, and actually performing such fabrication. As a

Fig. 3. Large throat depth welding station

flight quality fabrication shop, an air-conditioned "white room" environment is maintained because the output of welding power supplies drift with temperature sufficiently to require close temperature control. The facility has been equipped with three welding stations designed for cordwood module fabrication and one station for much larger work, such as matrix fabrication or repair (Fig. 3). The major difficulty in providing a flight-quality-level welding facility is the proper training of personnel, both for operation and technical direction, since resistance welding is not a widely known technology. To this end, a moderate amount of prototype fabrication has been performed and a very limited amount of flight hardware has been assembled.

Items fabricated ranged from the *Mariner C* (Mars Mission, 1964) Mars Gate and Absorbtivity Standard Thermometer Assembly (Fig. 4) to some moderately complex digital circuit modules for a prototype model of the nonreal-time, data automation systems (NRT DAS) from *Mariner C* assembled with typical electronic components (Fig. 5). The NRT DAS modules were also assembled into subassemblies in which portions of the interconnection were welded.

A follow-on effort to the NRT DAS prototype modules is a design study to provide an all-welded subsystem. Prototypes of this system will be fabricated and a complete mockup fabricated and subjected to vibration tests to determine the structural limitations of the projected design concept.

Fig. 4. *Mariner* C absorptivity standard — thermometer standard
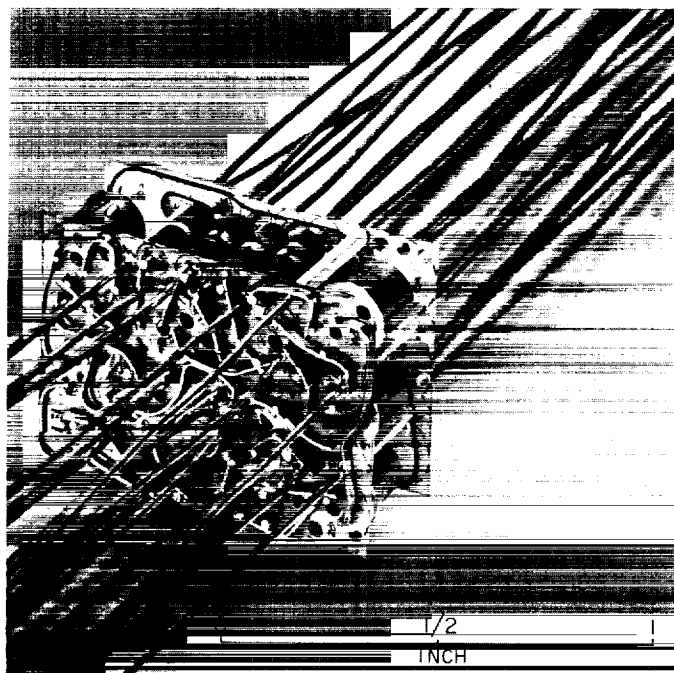
Fig. 5. Emitter follower module

ENGINEERING FACILITIES DIVISION

VII. Aerodynamic Facilities

A. Wind Tunnels

E. A. Nierengarten, P. Jaffe, G. M. Goranson, J. J. Minich,
R. W. Weaver, and M. J. Argoud

1. Mariner Blunt Body Stability,
E. A. Nierengarten and P. Jaffe

To obtain static aerodynamic coefficients a wind tunnel test was performed on *Mariner* planetary entry configurations; the test is outlined and the configurations are described in Ref. 1. Some results of the test at Mach number 4.54 are presented here, along with results as predicted by the unmodified Newtonian impact theory.

In particular, the normal force coefficient (C_N) for each configuration, as obtained from the test and theory, is presented as a function of the angle of attack. (At 90-deg angle of attack the normal force is parallel to the flow and is, therefore, the drag.) Fig. 1 shows the influence of centerbody length on the normal force coefficient for configurations having similar forebodies and afterbodies. Agreement of theory with actual C_N is reasonably good. The family of configurations considered in Fig. 2 has similar forebodies with various centerbody lengths and conical afterbody angles. The impact theory does not adequately describe the actual normal force coefficients for this family.
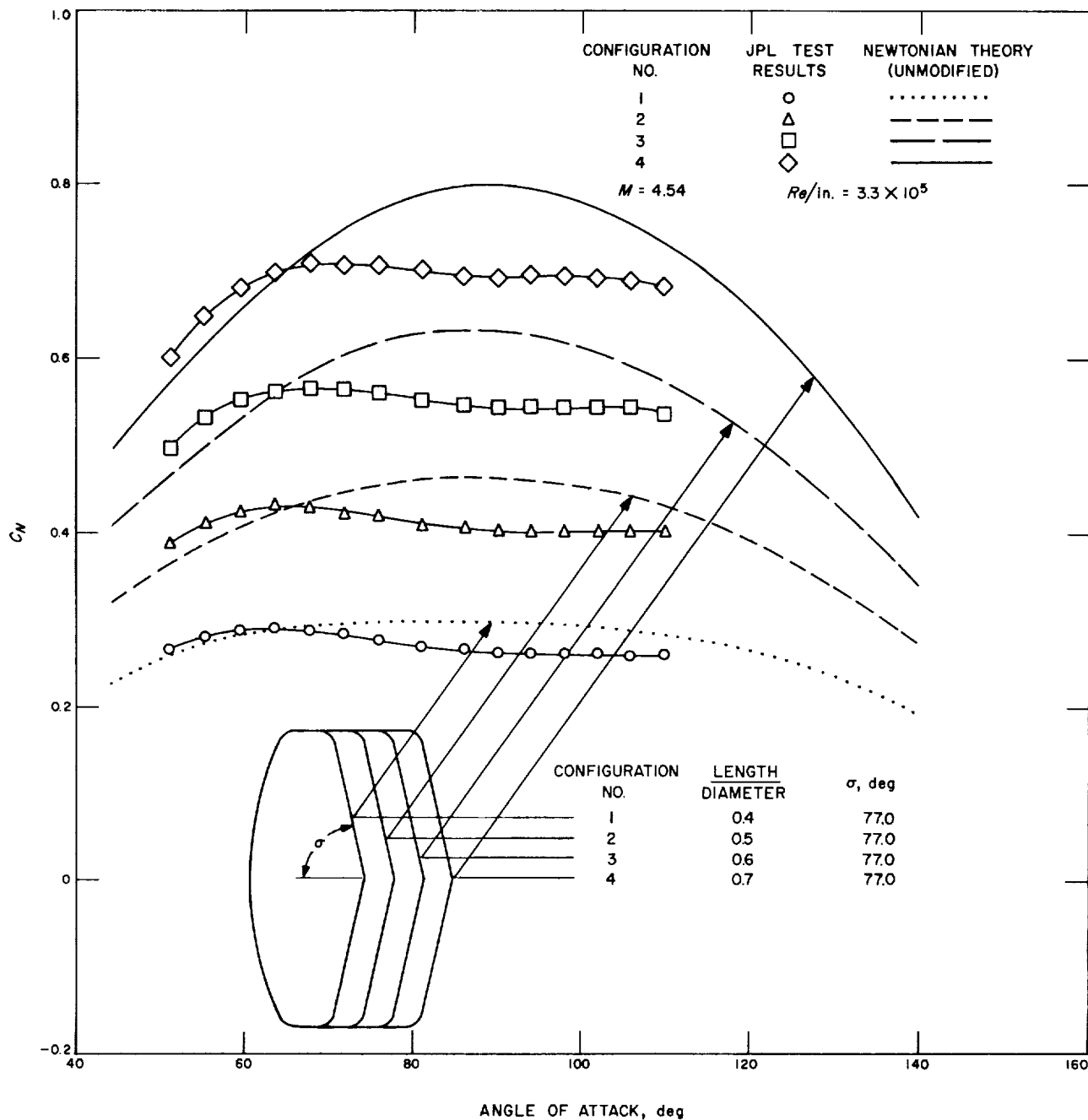
Fig. 1. Normal force coefficient (Configurations 1,2,3,4)

2. Free-Flight Body-Drogue Studies, G. M. Goranson

Wind Tunnel Test 20-579 was conducted as a preliminary investigation of supersonic forebody-drogue configurations in free flight with regard to both drag and dynamic behavior. A previous test (WT 20-567, Ref. 2) indicated that a drogue has a significant effect on the

stability of a forebody at high oscillation amplitudes. Subsequently, a more comprehensive investigation program was initiated, and Test 20-579 is an early result.

The following parameters are considered to be significant to the testing of the forebody-drogue system in free flight:

CONFIGURATION NO.	JPL TEST RESULTS	NEWTONIAN THEORY (UNMODIFIED)
5	O	··········
6	△	‒ ‒ ‒ ‒
7	□	—————
8	◇	—————

$M = 4.54$ $Re/\text{in.} = 3.3 \times 10^5$

CONFIGURATION NO.	$\dfrac{\text{LENGTH}}{\text{DIAMETER}}$	σ, deg
5	0.92	37.2
6	0.87	44.6
7	0.86	51.6
8	0.89	57.2
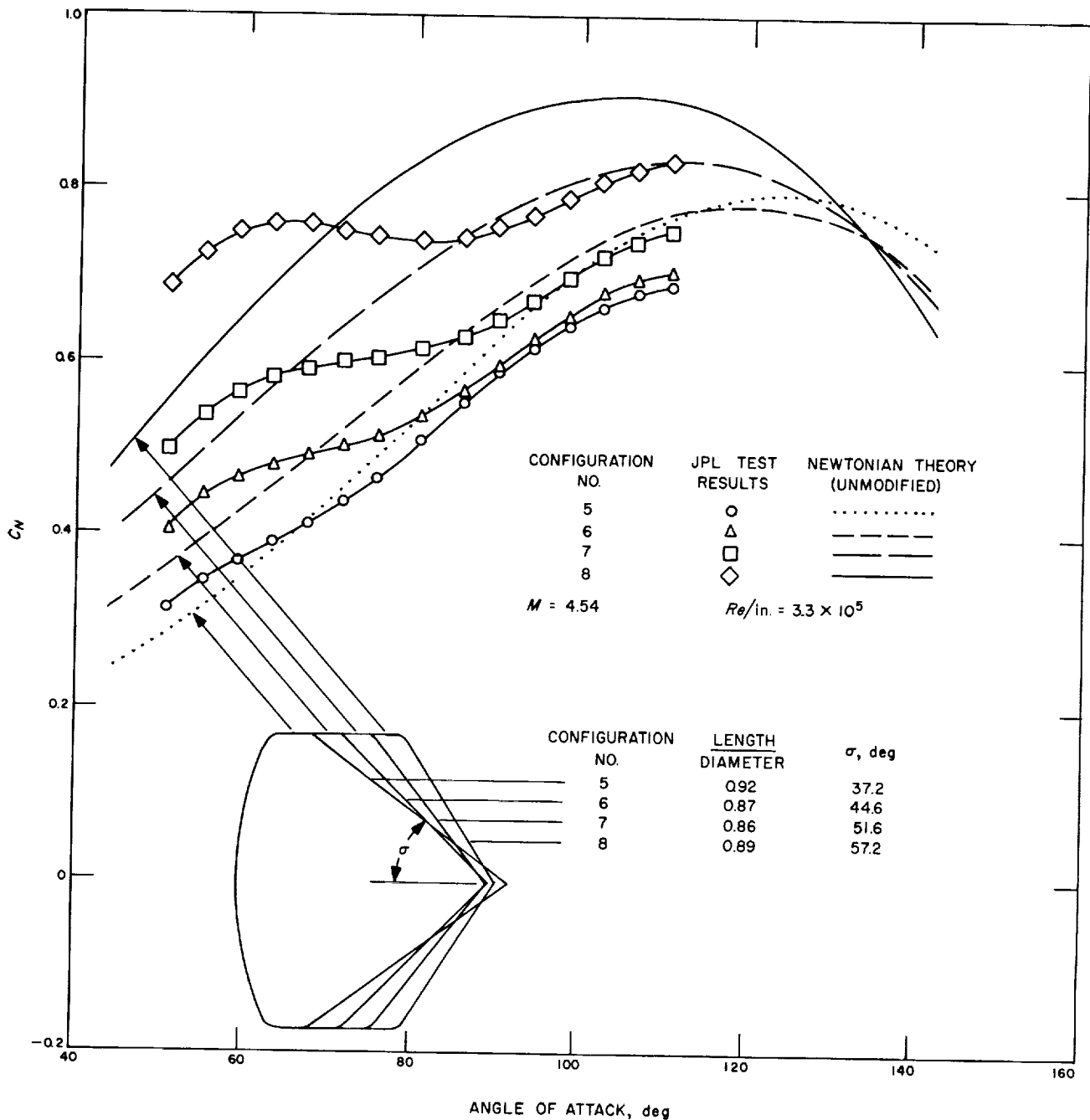
Fig. 2. Normal force coefficient (Configurations 5,6,7,8)

Geometric
(1) Forebody shape.
(2) Drogue shape.
(3) Ratio, drogue diameter to forebody diameter (d/D).
(4) Ratio, tow line length to forebody diameter (l/D).
(5) Attachment method of tow line to the bodies.
(6) Launch angle of attack of forebody.

(7) Launch mode (i.e, relative position of drogue and forebody at launch).

Dynamic
(8) Ratio, mass of forebody to mass of drogue.
(9) Nature of tow line material (i.e., mass, flexibility, elasticity).

(10) Moments of inertia of each body.

Flow

(11) Mach number.

(12) Reynolds number.

Figs. 3 and 4 illustrate the ranges of geometric parameters 1 through 5 used in this test. Launch angle of attack of forebody was 0, 35, 90, or 180 deg. Launch mode refers to the two alternative methods used for releasing the models into free flight: either restraining only the forebody prior to release (with the drogue trailing in the stream), or holding both bodies side by side prior to release thus allowing the drogue to "deploy" downstream to the length of the tow line, after release into free flight. No attempt was made to vary the remaining parameters for this test. The mass of the drogues was simply made small compared to the forebodies; the tow line was nylon

fishing line (10-lb test); forebody moment of inertia/mass ratio was designed to yield the maximum number of oscillations in the test section. Test conditions were: Mach number, 4.5; Reynolds number/in., 7.5×10^4; dynamic pressure, 0.68 psia.

Definitive quantitative results of this test will be presented when the data are reduced and analyzed. However, some preliminary qualitative observations are: drogue drag values for the 0-deg launch angle cases are significantly below free-stream values and show a definite *decrease* in drag with increasing trailing distances for the range tested. Significant pitch damping of the forebody oscillation occurs in both the 90- and 35-deg launch angle regimes. Damping has been observed, even in the case of an apparently slack tow line, throughout the flight.

Results of this test thus far indicate a large amount of information of potentially great interest to be gained through the investigation of the above parameters. Further analysis and investigation are planned.

3. Gemini Abort System Dynamic Stability, J. J. Minich

Wind Tunnel Test 20-583 was a NASA-sponsored test of a Manned Space Flight Center (Houston, Texas) scale model of the *Gemini* abort seat and passenger. Purpose of the test was to obtain dynamic stability data for the seat and passenger at abort conditions of Mach number and dynamic pressure. The approximate aerodynamic
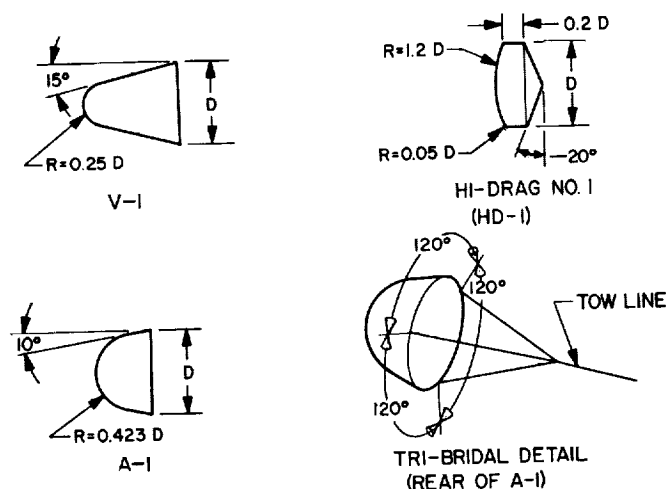
Fig. 3. Forebody configurations
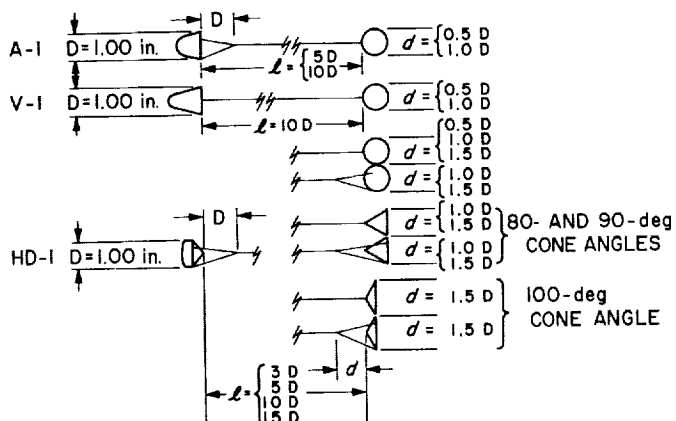
Fig. 4. Forebody-drogue combinations

Fig. 5. Dynamic stability models
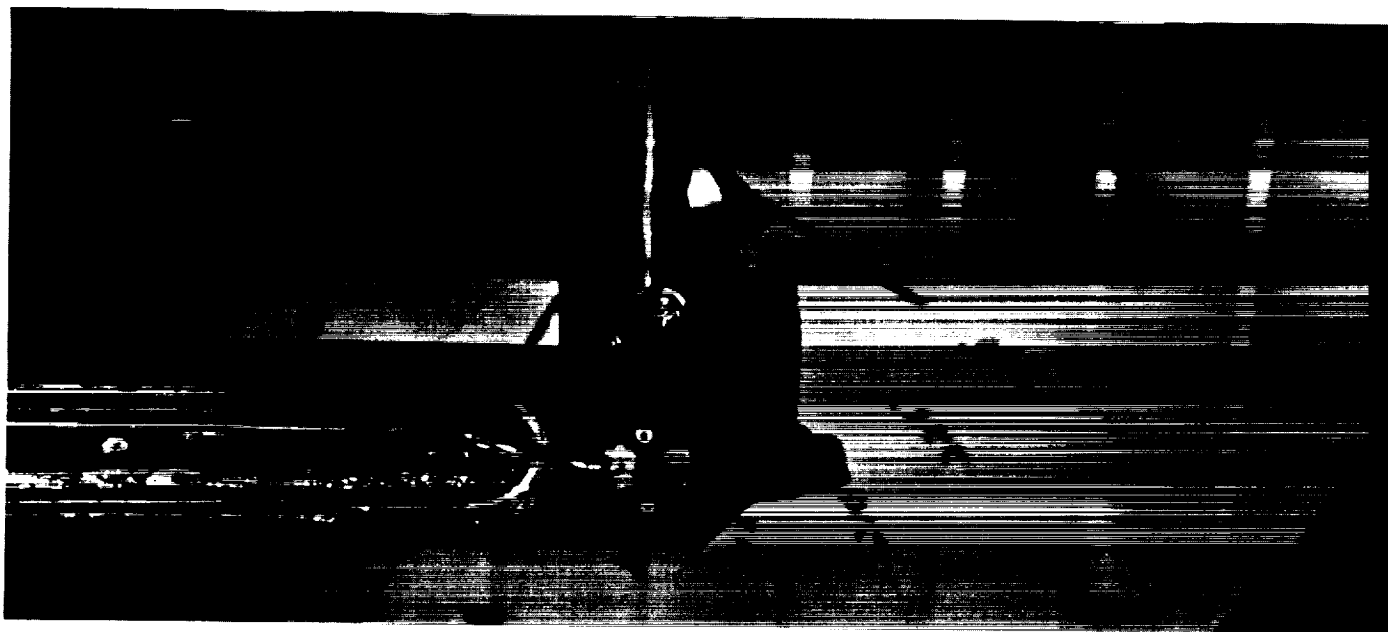
Fig. 6. Dynamic stability test installation

parameters for the test were Mach numbers of 1.33, 1.65, 2.21, and 2.81 and corresponding Reynolds numbers/in. of 0.06 × 10⁶ through 0.16 × 10⁶. The test variables and ranges were release angle of attack from 0 to 340 deg. The models were supported on a transverse rod with ball bearings. Oscillatory motion data were obtained with a conventional 16 mm movie camera. Model configurations, along with the test installation, are shown in Figs. 5 and 6.

4. Advanced X-15 Stability, R. W. Weaver

A series of NASA-sponsored Wind Tunnel Tests (20-582 and 21-160) was recently completed on the modified version of the extended performance X-15 research aircraft. The configurations tested to obtain stability and control data were essentially the same as those in Ref. 3. The most significant modification tested was the addition of canard control surfaces on the forward fuselage (Fig. 7). The canards were added to give longitudinal
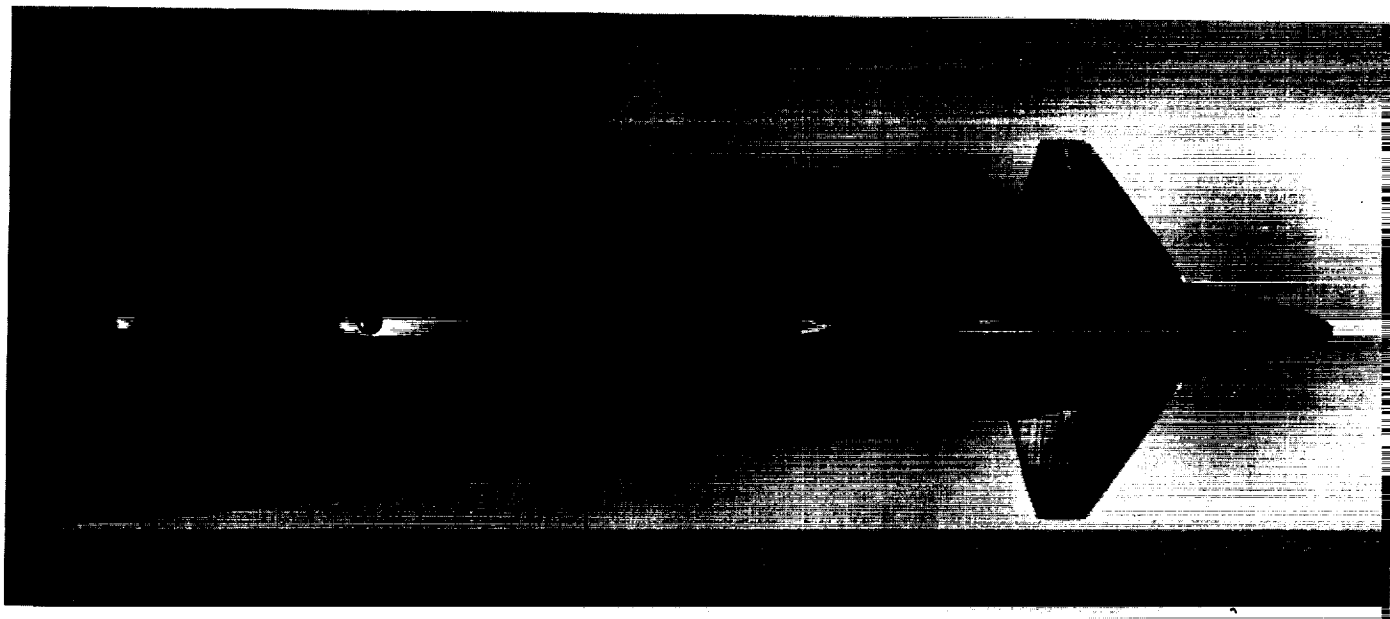
Fig. 7. Canard control surfaces on X-15 model

control during high angle-of-attack re-entry flights. The approximate aerodynamic parameters for the test were Mach numbers from 1.48 to 8.00 and corresponding Reynolds numbers/in. from 0.35×10^6 to 0.21×10^6. The test variables and ranges were angle of attack from -10 to $+44$ deg, and angle of yaw from -4 to $+10$ deg.
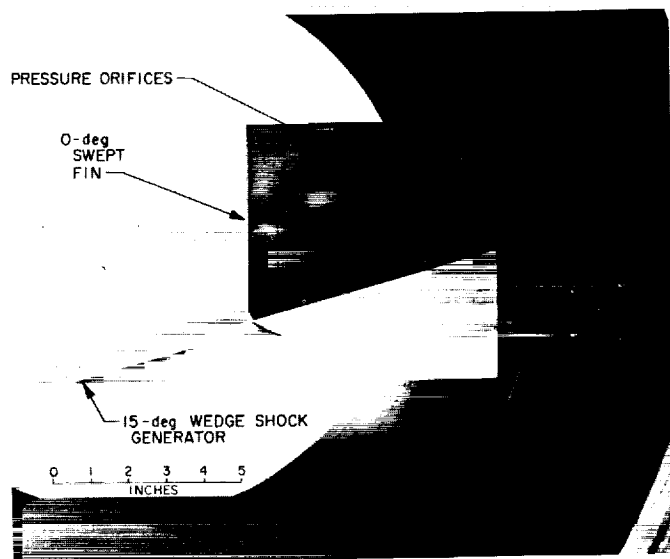
Fig. 8. Pressure wedge model installation

5. Saturn Fin Shock Impingement Heat Transfer,
J. J. Minich

Wind Tunnel Test 21-155A was of the Boeing Company (Huntsville, Alabama) shock impingement heat transfer model. The NASA-sponsored test was made to obtain heating rates and pressure distributions in the area of shock-wave impingement. Approximate aerodynamic parameters for the test were Mach numbers 5.0, 6.0, and 8.0 with corresponding Reynolds numbers/in. of 2 through 10×10^4. The test variables and ranges were fin yaw angle from 0 to 10 deg. The model configuration consisted of a 15-deg wedge as a shock-wave generator and an instrumented, 0-deg swept fin (Fig. 8). The test was a continuation of WT 21-155 (Ref. 4) in which heating studies were conducted for a 30-deg swept fin with the same shock generator and at the same tunnel conditions.

6. Aerobee Stability, *M. J. Argoud*

NASA-sponsored Wind Tunnel Test 20-587 was of the Space General Corporation (El Monte, California) 0.05909-scale *Aerobee* 350 wind tunnel model. The test was conducted to obtain aerodynamic force and moment coefficients as a function of roll position. Approximate
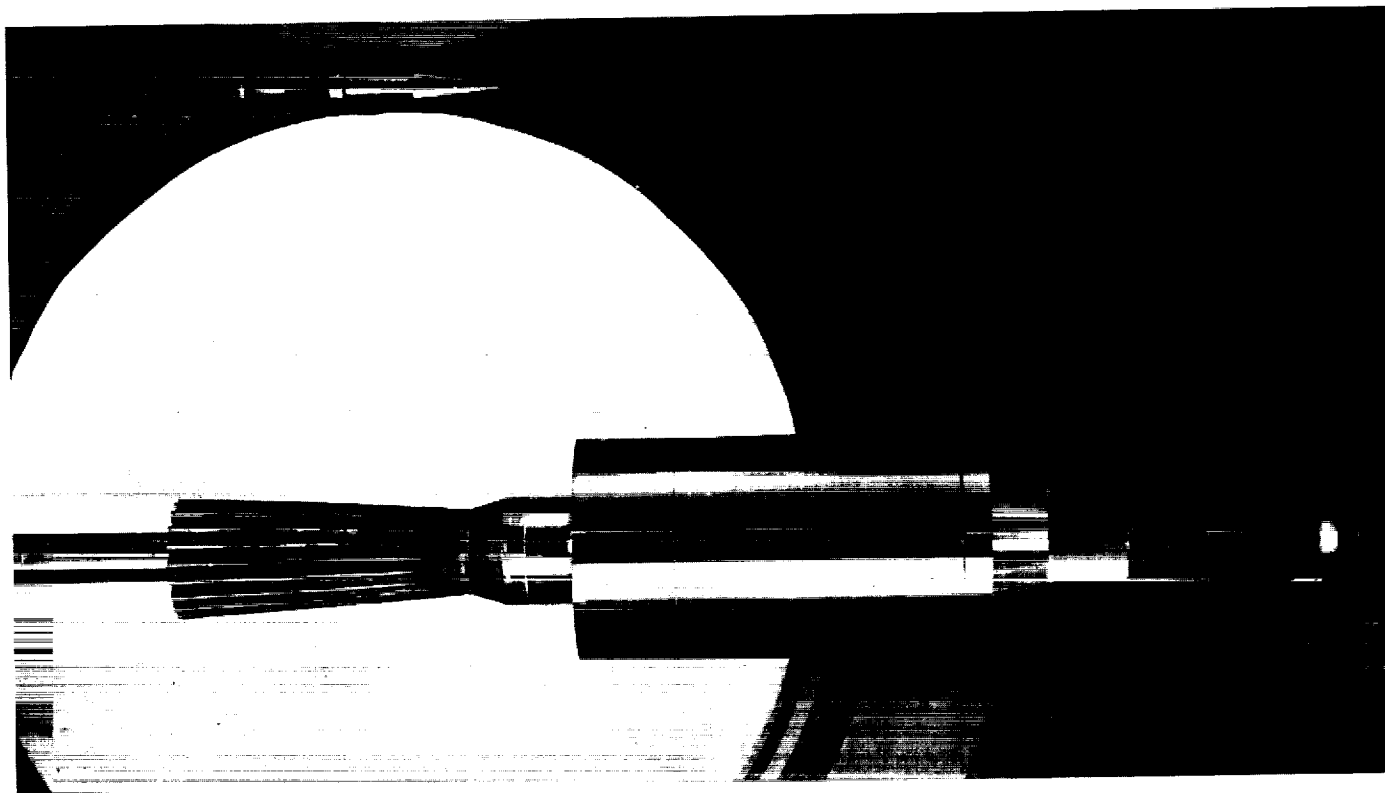
Fig. 9. Sting-mounted installation of Configuration T in the 21-in. HWT

aerodynamic parameters were Mach numbers 2.8 to 5.0 and corresponding Reynolds numbers/in. of 0.34×10^6 to 0.17×10^6. The test variables and ranges were angle of attack from -4 to $+16$ deg, and angle of roll from -10 to $+90$ deg. The model configuration was made up of a tangent ogive nose with a fineness ratio of five, followed by a cylindrical afterbody with cruciform tail surfaces at the aft end. Four semicircular shrouds were spaced about the periphery of the vehicle beginning at the aft end of the ogive and ending at the leading edges of the tail fins. Forces and moments were obtained for the complete configuration.

7. Expandable Rocket Nozzles, R. W. Weaver

Tests were recently conducted in the JPL Hypersonic Wind Tunnel (HWT) on expandable rocket nozzles for upper stage vehicles. The test program was initiated by Aerojet-General Corporation (Sacramento, California) and is sponsored by the Department of Defense. Objective of the tests was to investigate the aerodynamic effects of external flow on the opening characteristics of the nozzles.

Three models were tested. Each was a possible upper stage vehicle. The models were mounted in one of two ways: conventional sting mount (Fig. 9), or floor-mounted strut (Fig. 10). The expandable portion of the nozzles (referred to as skirts) was made using two different methods: (1) electroforming the skirts on male molds which produced skirts that were somewhat stiffer than desired and (2) mechanically forming the convolutions in the skirt which yielded skirts with the scaled characteristics of the full-scale items.

Fig. 10. Strut-mounted installation of Configuration V in the 21-in. HWT

The rocket gas flow was simulated by using gaseous nitrogen. A problem arose because of the relative mass flows between the wind tunnel and the rocket. Under certain conditions the mass flow from the rocket was approximately equal to the mass flow of the tunnel. The result was separated flow in the test section which gave erroneous data for those particular runs. The average aerodynamic parameters for these tests were Mach numbers from 6.5 to 9.0 and Reynolds numbers/in. from 0.02×10^6 to 0.21×10^6.

B. Hypervelocity Laboratory

F. R. Livingston, G. M. Thomas, and W. A. Menard

1. H_2-N_2 Shock-Tube Results, F. R. Livingston

Unheated hydrogen has recently been used as the driver gas in an existing 3-in.-D, 17-ft-long shock tube (Fig. 11) in the Hypervelocity Laboratory in order to obtain shock velocities of over 12,000 ft/sec in nitrogen gas. To reach the maximum test gas temperature in an unheated shock tube, hydrogen gas must be used to drive the shock. The theoretical shock velocity limit using hydrogen driver gas is more than twice that of helium, the next best driver gas.

Performance was computed for the 3-in. shock tube by methods of shock-tube theory accounting for the steady flow in the 2.125-in.-square orifice at the diaphragm station. The hydrogen driver gas was assumed to be a perfect gas. The nitrogen gas incident and reflected shock conditions were obtained from the real gas computations of Ref. 5. Experimental shock speed results were obtained by measuring the time of arrival of the incident shock at stations near the driven end of the shock tube. The experimental shock speed-pressure ratio results are shown in Fig. 12 along with theoretical results. The experimental shock speed is somewhat less than expected from theory. Loss in speed is usually attributed to viscous effects not included in the theoretical considerations. However, viscous attenuation of the shock is certainly not noticeable in Fig. 13 where shock position is shown as a function of time. Additional investigation is required here. In addition to shock speed, the pressure behind the incident and reflected shock was measured with a Kistler, Model 601, piezoelectric pressure transducer located at Station 6. Reflected shock
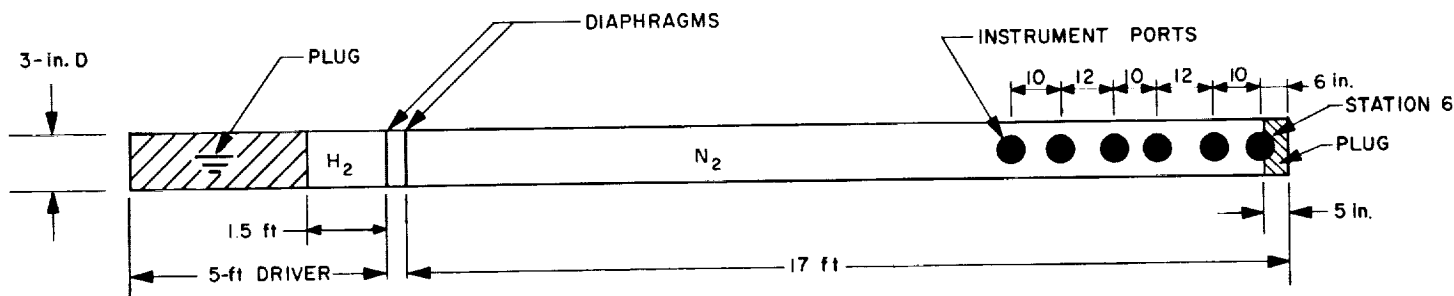
Fig. 11. Schematic diagram of 3-in. shock tube
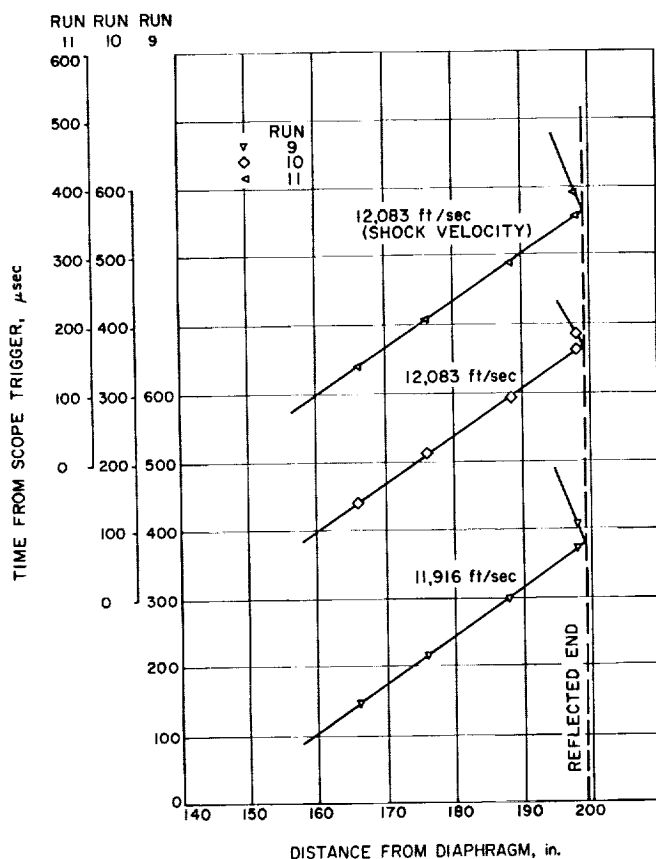
Fig. 12. Variation of shock speed with diaphragm pressure ratio

SHOCK VELOCITY U_S, × 10⁻³ ft/sec

Fig. 13. Variation of incident and reflected shock position with time

pressure ratios show good agreement with the calculations of Ref. 5.

Safety precautions taken while using hydrogen include ventilation of the overhead regions in the building and the venting of used gases to the outside of the building. Tests show that hydrogen can be safely used to drive shocks at speeds to 12,000 ft/sec in nitrogen. Experimental pressure-velocity relationships show reasonable agreement with theoretical calculations.

2. Planetary Gas Radiation, G. M. Thomas and W. A. Menard

An experimental investigation is currently being conducted in the electric shock tube in order to: (1) measure the equilibrium and nonequilibrium spectral radiation
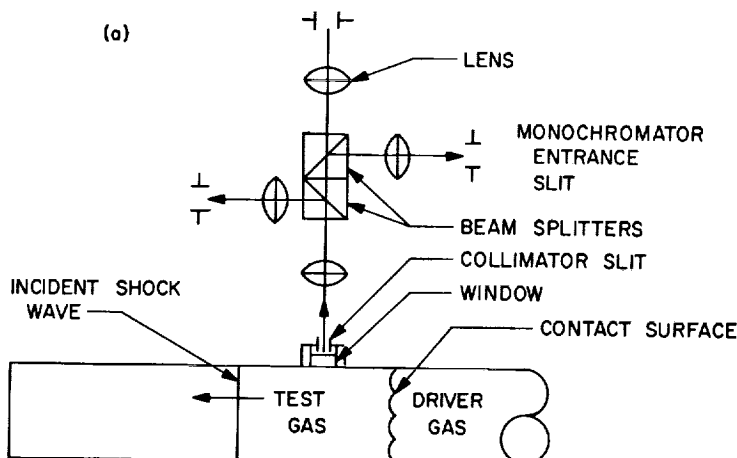
Fig. 14. (a) Schematic of optical path; (b) apparatus for making spectral radiation measurements

behind an incident shock wave and (2) measure the total radiative heat transfer to the stagnation point of a model in gas mixtures similar to the atmospheres of Mars and Venus. The atmospheres of both planets are believed to be mainly carbon dioxide and nitrogen with possible

(a) 9% CO_2 – 91% N_2 UPPER TRACE – λ = 6390 – 6610 Å
U_S = 24,600 ft/sec   LOWER TRACE – λ = 4705 – 4795 Å
P_1 = 0.250 mm Hg
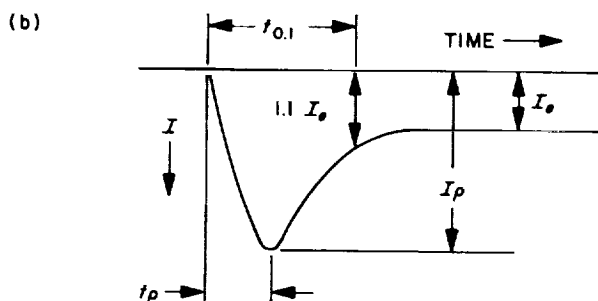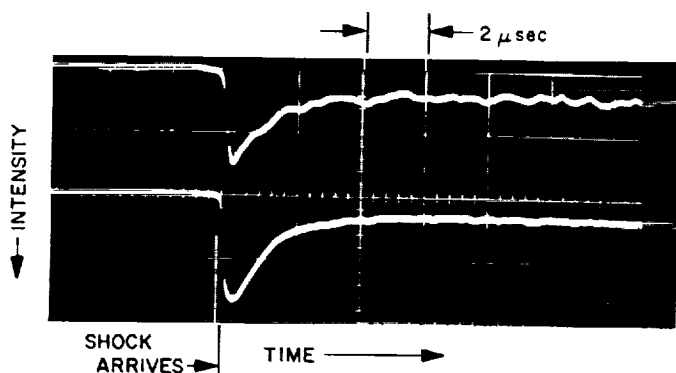
Fig. 15. (a) Oscilloscope traces from two monochromators viewing the test gas radiation; (b) measurements obtained from each trace

trace amounts of argon, but the exact percentages of the components are not known. For this reason it is necessary to investigate the radiation from several different mixtures to determine the effect of composition on the radiation.

In the current study the spectral radiation behind the incident shock wave and the radiative heat transfer to the stagnation point of a flat-faced cylinder are being measured for shock speeds ranging from 20,000 to 30,000 ft/sec, and gas mixtures of 9, 30, and 100% carbon dioxide with the remainder being nitrogen. The initial driven tube pressure is 0.250 mm Hg. Spectral radiation is measured over the wavelength region from 3600 to 10,000 A and the total radiation measurements cover the region from 0.3 to 2.7 μ.

The spectral radiation behind the incident shock wave is measured by three Perkin-Elmer monochromators (Ref. 6) as shown schematically in Fig. 14. A typical oscillogram of the radiation intensity from two monochromator channels is presented in Fig. 15. The various measurements which are made from these traces are indicated on the figure. For the nonequilibrium region the following measurements are made: peak intensity, time to peak (t_p), and time to reach 1.1 times the equilibrium intensity $(t_{0.1})$. The nonequilibrium integrated intensity directed toward a body from a slab of gas of thickness $U_s t_{0.1}$ is also calculated. Equilibrium radiation intensity is also obtained from a different oscillogram in which the scope gain is adjusted for improved accuracy in measuring the trace deflection due to the equilibrium radiation. An example of a spectrum is given in Fig. 16.

A complete description of the total radiation gage has been presented previously (Ref. 6).
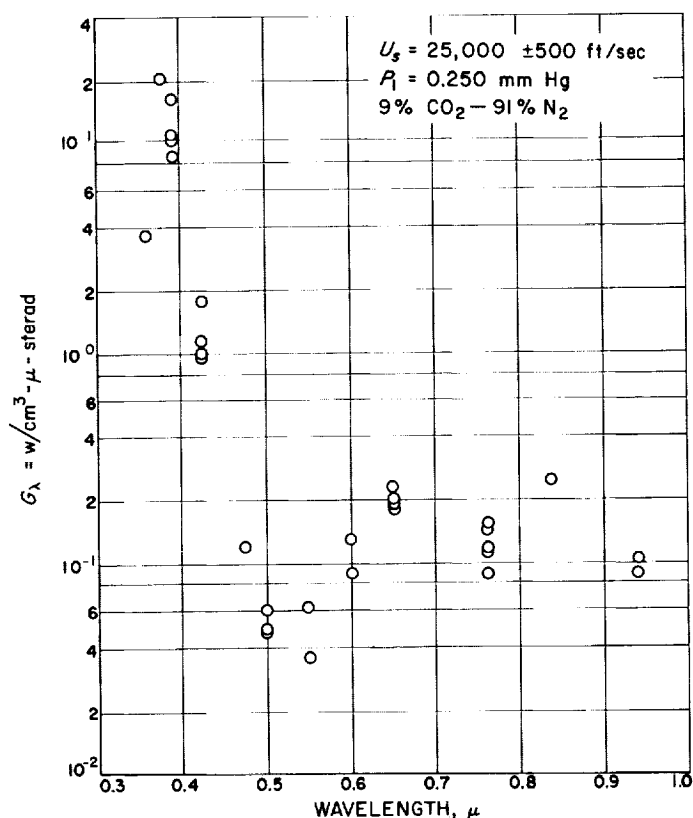
Fig. 16. Equilibrium spectral radiation

References

1. Nierengarten, E. A., and Jaffe, P., "Mariner Configuration Study," Space Programs Summary No. 37-25, Vol. IV, pp. 49–51, February 29, 1964.

2. Goranson, G. M., "Mariner B Studies," Space Programs Summary No. 37-24, Vol. IV, pp. 70–72, December 31, 1963.

3. Weaver, R. W., "X-15," Space Programs Summary No. 37-23, Vol. IV, p. 65, October 31, 1963.

4. Minich, J. J., "Saturn Shock-Wave Impingement," Space Programs Summary No. 37-24, Vol. IV, pp. 68–69, December 31, 1963.

5. Ahtye, Warren F., and Peng, Tzy-Cheng, "Approximations for the Thermodynamic and Transport Properties of High-Temperature Nitrogen with Shock Tube Applications," NASA TN D-1303, National Aeronautics and Space Administration, Washington, D. C., July 1962.

6. Thomas, G. M., "Experimental Radiation Measurement Techniques," Space Programs Summary No. 37-24, Vol. IV, pp. 78–83, December 31, 1963.

PROPULSION DIVISION

VIII. Solid Propellant Engineering

A. Low-Pressure Unstable Combustion Analysis

L. Strand

1. Mechanism of Low-Pressure Combustion Extinction

In Ref. 1 the solid rocket characteristic chamber length L^*, equal to the ratio of the motor chamber free volume to the nozzle throat area, was shown to correlate with the mean critical chamber pressure $\bar{p}_{cr}$ for conditional combustion stability for a given solid propellant. The following expression was derived:

$$L^* = \left[\frac{4 \alpha C_D RT_f (\tau_n)_{cr}}{M a^2 g}\right] \bar{p}_{cr}^{-2n} \qquad (1)$$

where $(\tau_n)_{cr}$ is the critical rocket chamber relaxation time constant non-dimensionalized by $4 \alpha/\bar{r}^2$, and the other terms are defined in Table 1. According to the model presented, at or below the pressure $\bar{p}_{cr}$ a random pressure

perturbation would be reinforced by the combustion mechanism, resulting in oscillatory combustion of increasing amplitude. Over the pressure region of instability the bracketed terms can be treated as constant. Therefore, for a given propellant a log-log plot of L^* versus $\bar{p}_{cr}$ should be a straight line with slope equal to $-2n$.

In the experimental investigation of low-pressure combustion limits of several JPL solid propellants (Ref. 2), it was found that the mean chamber pressure prior to extinction of combustion was correlated by the L^* parameter value at extinction on a log-log plot. In a majority of the 3-in. internal diameter test motor firings, low-frequency pressure oscillations could be detected on the recorder pressure trace occurring prior to extinction. This led to the belief that these low-frequency oscillations in pressure are a manifestation of the combustion instability which results in extinguishment of combustion (Ref. 3).

The mechanism of extinction that precludes further combustion from being possible as the chamber pressure decreases for a regressive burning solid propellant grain

Table 1. Definition of terms

A_b	burning area
A_p	port area
A_t	nozzle throat area
a	constant in de Saint-Robert's Law
C_1, C_2	parameters defined in Eq. (27)
C_D	discharge coefficient
c_p	propellant constant pressure specific heat
f	heat flux to propellant surface
G	port mass velocity
g	gravitation constant
k	propellant thermal conductivity
L	distance along grain
L^*	characteristic chamber length
M	gaseous molecular weight
n	pressure exponent in de Saint-Robert's Law
p	pressure
$\bar{p}_e$	chamber pressure at combustion extinction
$\bar{p}^*$	critical chamber pressure defined by Eq. (36)
Q	heat of gasification (positive value endothermic)
q	pre-heat energy
R	universal gas constant
r	burning rate
r_1	burning rate defined in Eq. (25)
r_e	erosive burning rate component
$\bar{r}_p$	linear burning rate component
T	temperature
T_f	flame temperature
T_o	initial propellant temperature
T_s	surface temperature
t	time
V	chamber volume
x	dimension normal to propellant surface
α	thermal diffusivity
β, γ	constants defined in Eq. (43)
ρ	propellant density
τ	thermal characteristic time
τ_n	chamber relaxation time, nondimensionalized by $4\alpha/\bar{r}^2$

Superscript
- steady state

Subscripts
cr	critical
s	surface

is not clearly understood. The postulated triggering mechanism for low-frequency instability, the lag in propellant burning rate response (due to the solid-phase thermal gradient) to a random pressure perturbation, seems unlikely because it can be shown that there should be no transient burning rate response to the very low-frequency pressure oscillations (Ref. 4). Also, from Ref. 2 test motors fired under the same conditions would extinguish at approximately the same mean chamber pressure and L^*, but for some of the tests little or no combustion instability could be observed on the oscillograph pressure traces prior to extinction. In a paper by G. von Elbe (Ref. 5), a model for non-steady combustion during a change in heat flux to the propellant surface and a criterion for combustion extinction are postulated that appear applicable to the extinction of the regressively burning propellant charges motor fired at JPL. The following is a brief summary of this model and an extension of the theory into a familiar form.

2. Non-Steady Combustion Model

The model consists of a stationary propellant burning surface with reactants being transported from their initial temperature T_o to the surface temperature T_s, undergoing gasification, and reacting in the flame zone above the propellant surface, the heat of gasification being neglected in this analysis. The propellant regression rate r is assumed to adjust immediately to any change in heat flux to the propellant surface.

Fig. 1 illustrates how the combustion model adjusts to a reduction in heat flux to the propellant surface from f to $f-df$ due, in this analysis, to a reduction in chamber pressure. The integrated energy per unit surface area
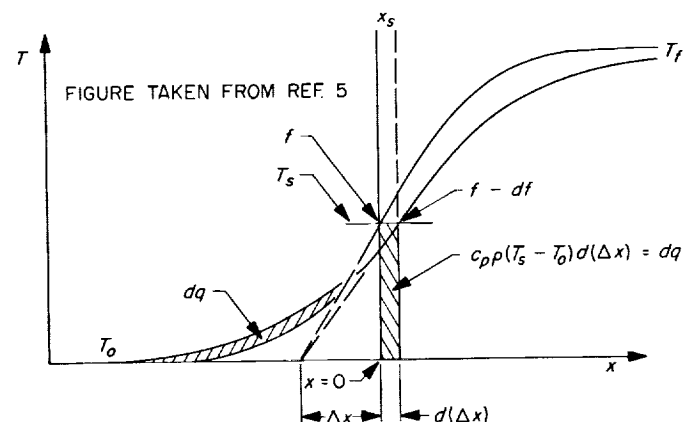
Fig. 1. Model for non-steady solid propellant combustion

under the propellant solid phase thermal gradient is called the pre-heat energy and is designated q. In terms of the model, q can be pictured as the energy necessary to raise the propellant reactants' temperature from ambient temperature T_o to the gasification temperature T_s. A decrease in heat flux to the propellant surface results in two changes: (1) the temperature at the reference plane ($x = 0$) drops below T_s, so that the reference plane moves forward to the T_x position on the new temperature gradient; and (2) a portion $d(q)$ of the heat q stored in the propellant is conducted across the old temperature gradient and establishes a new gradient of lower slope. Let f be the instantaneous heat flux and r the instantaneous burning rate during a period of change. Then, in a time interval dt, fdt heat per unit surface area flows by conduction into the propellant surface and $fdt-dq$ energy is convected back to the propellant surface by the propellant reactants. This latter energy consists of the energy $c_p \rho (T_s - T_o) rdt$ plus $c_p \rho (T_s - T_o) d(\Delta x)$, where $d(\Delta x)$ is the distance of movement of the x_s-surface.

Putting in equation form,

$$c_p \rho (T_s - T_o) \left[rdt + d(\Delta x) \right] = fdt - dq \qquad (2)[1]$$

or

$$c_p \rho (T_s - T_o) (r + d(\Delta x)/dt) = f - dq/dt \qquad (3)[1]$$

In Ref. 5 the rate of change of heat flux is considered to be sufficiently slow so that the time of adjustment of the temperature gradient (the thermal relaxation time) $d\tau$ is small compared to the time for heat flux to change from f to $f-df$. Therefore, equations for steady-state combustion were employed in obtaining time derivatives of the various parameters by differentiating these equations with respect to f and multiplying the derivatives by the rate of change of heat flux df/dt. These steady-state equations are derived below for completeness.

3. Equations for Steady-State Combustion

At any distance below the propellant grain surface a heat balance yields

$$c_p \rho (T - T_o) r = kdT/dx \qquad (4)[1]$$

At the propellant surface, but in the solid phase,

$$c_p \rho (T_s - T_o) r = k (dT/dx)_{x=0}$$

$$= f \qquad (5)[1]$$

To define a Δx,

$$\frac{T_s - T_o}{\Delta x} = \left(\frac{dT}{dx} \right)_{x=0} \qquad (6)[1]$$

then

$$\Delta x = \frac{T_s - T_o}{(dT/dx)_{x=0}} \qquad (7)$$

Substituting Eq. (5) into Eq. (7),

$$\Delta x = \frac{k}{c_p \rho r} = \frac{\alpha}{r} \qquad (8)[1]$$

From Fig. 2 and the definition of the pre-heat q,

$$dq = c_p \rho (T - T_o) dx \qquad (9)[1]$$

Inserting Eq. (4),

$$dq = \frac{k}{r} \frac{dT}{dx} \cdot dx$$

$$= \frac{k}{r} dT \qquad (10)$$

Integrating Eq. (10) along the solid-phase temperature profile,

$$\int dq = \frac{k}{r} \int_{T_o}^{T_s} dT$$

$$q = \frac{k}{r} (T_s - T_o) \qquad (11)[1]$$

Combining Eqs. (8) and (11),

$$q = \rho c_p (T_s - T_o) \Delta x \qquad (12)[1]$$

The widely used characteristic time τ is defined by

$$\tau = \frac{\Delta x}{r} = \frac{\alpha}{r^2} = \frac{k}{c_p \rho r^2} \qquad (13)[1]$$

During the time τ the quantity of heat τf crosses the grain surface. From Eqs. (5) and (13),

$$\tau f = \frac{k}{c_p \rho r^2} c_p \rho (T_s - T_o) r$$

$$= \frac{k}{r} (T_s - T_o)$$

Then, from Eq. (11),

$$\tau f = q \qquad (14)[1]$$

The product fq gives

$$fq = f^2\tau$$

From Eq. (5) and Eq. (13)

$$f^2\tau = c_p \rho k (T_s - T_o)^2 \qquad (15)^1$$

which is a constant.

4. Non-Steady Combustion Ballistics Equations

Returning to the non-steady problem, differentiating Eq. (12) with respect to time gives

$$c_p \rho (T_s - T_o) d(\Delta x)/dt = dq/dt$$

Incorporating in Eq. (3),

$$c_p \rho (T_s - T_o) r = f - 2 dq/dt \qquad (16)^1$$

Quoting Ref. 5, "When $f = 2 dq/dt$, r is zero. This means that the gasification of the propellant material ceases and the zone of chemical reaction disappears. When $f \leq 2 dq/dt$, the surface temperature drops below T_s." In other words, extinction occurs when the decreasing heat flux to the propellant surface is not able to maintain the propellant gasifying rate plus the solid-phase pre-heat energy rate of increase.

From Eq. (14),

$$\frac{dq}{dt} = \tau \frac{df}{dt} + f \frac{d\tau}{dt} \qquad (17)$$

Differentiating Eq. (15) with respect to time,

$$\tau (2f) \frac{df}{dt} + f^2 \frac{d\tau}{dt} = 0 \qquad (18)$$

Eliminating df/dt by combining Eqs. (17) and (18),

$$2 \frac{dq}{dt} = f \frac{d\tau}{dt} \qquad (19)^1$$

Substituting Eq. (19) into Eq. (16),

$$c_p \rho (T_s - T_o) r = f \left(1 - \frac{d\tau}{dt} \right) \qquad (20)^1$$

Rewriting Eq. (15),

$$\tau = \frac{c_p \rho k}{f^2} (T_s - T_o)^2$$

Differentiating with respect to time,

$$\frac{d\tau}{dt} = - \frac{2 c_p \rho k}{f^3} (T_s - T_o)^2 \frac{df}{dt} \qquad (21)$$

Considering only the change of flux with pressure,

$$\frac{df}{dt} = \frac{\partial f}{\partial p} \frac{dp}{dt} \qquad (22)^1$$

or Eq. (21) becomes

$$\frac{d\tau}{dt} = - \frac{2 c_p \rho k}{f^3} (T_s - T_o)^2 \frac{\partial f}{\partial p} \frac{dp}{dt} \qquad (23)$$

Inserting in Eq. (20),

$$c_p \rho (T_s - T_o) r = f \left[1 + \frac{2 c_p \rho k}{f^3} (T_s - T_o)^2 \frac{\partial f}{\partial p} \frac{dp}{dt} \right] \qquad (24)^1$$

As a major simplification, von Elbe assigns to the heat flux existing at any instance a steady-state burning rate r_1 so that

$$c_p \rho (T_s - T_o) r_1 = f$$

and

$$\frac{\partial f}{\partial p} = c_p \rho (T_s - T_o) \frac{\partial r_1}{\partial p} \qquad (25)^1$$

Substituting in Eq. (24),

$$r = r_1 \left(1 + \frac{2k}{c_p \rho} \frac{1}{r_1^3} \frac{\partial r_1}{\partial p} \frac{dp}{dt} \right) \qquad (26)^1$$

This equation gives the dependence of the burning rate on dp/dt.

The ballistic equation describing the rate of pressure change in a rocket motor is:

$$\frac{dp}{dt} = \left(\frac{A_b \rho R T_f}{VM} \right) r - \left(\frac{A_t C_D R T_f}{VMg} \right) p$$
$$= C_1 r - C_2 p \qquad (27)^1$$

where

$$C_1 = \frac{A_b \rho R T_f}{VM}$$

and

$$C_2 = \frac{A_t C_D R T_f}{VMg}$$

By inserting Eq. (26) for r in Eq. (27), a complete ballistic equation is obtained that includes the dependence of r on dp/dt. Thus,

$$\frac{dp}{dt} = \frac{C_1 r_1 - C_2 p}{1 - C_1 \frac{2k}{c_p \rho} \frac{1}{r_1^2} \frac{\partial r_1}{\partial p}} \qquad (28)^1$$

Substituting Eq. (28) for dp/dt in Eq. (26),

$$r = r_1 \left[1 + \frac{2k}{c_p \rho} \frac{1}{r_1^3} \frac{\partial r_1}{\partial p} \left(\frac{C_1 r_1 - C_2 p}{1 - C_1 \frac{2k}{c_p \rho} \frac{1}{r_1^2} \frac{\partial r_1}{\partial p}} \right) \right] \quad (29)$$

5. Extinction of Combustion

Returning to extinction criterion, extinction occurs when the regression or gasification rate equals zero. Setting the right side of Eq. (29) equal to zero,

$$1 + \frac{2k}{c_p \rho} \frac{1}{r_1^3} \frac{\partial r_1}{\partial p} \left(\frac{C_1 r_1 - C_2 p_e}{1 - C_1 \frac{2k}{c_p \rho} \frac{1}{r_1^2} \frac{\partial r_1}{\partial p}} \right) = 0 \quad (30)$$

When dp/dt is zero, the burning becomes its steady-state value $\bar{r}$. As the transition in heat flux (rate of pressure change) occurs more and more slowly, r_1 approaches the steady-state value $\bar{r}$. This approximation is used in Ref. 5. Also, de Saint-Robert's Law for the linear burning rate, $\bar{r}_p = a\bar{p}^n$, is used for $\bar{r}$, neglecting the erosive burning rate component, so that $\partial r_1/\partial p$ in Eq. (30) equals $\partial \bar{r}_p/\partial p = an\bar{p}^{n-1}$.

Making these substitutions in Eq. (30) and simplifying,

$$1 + \frac{\frac{2anC_1}{a\bar{p}_e^{n+1}} - \frac{2anC_2}{a^2\bar{p}_e^{2n}}}{1 - \frac{2anC_1}{a\bar{p}_e^{n+1}}} = 0 \quad (31)$$

where $\bar{p}_e$ is the quasi-steady-state chamber pressure at combustion extinction. Multiplying through by the denominator and cancelling like terms, Eq. (31) becomes

$$\frac{1 - \frac{2anC_2}{a^2\bar{p}_e^{2n}}}{1 - \frac{2anC_1}{a\bar{p}_e^{n+1}}} = 0 \quad (32)$$

For Eq. (32) to be true the numerator must equal zero or

$$\frac{2anC_2}{a^2\bar{p}_e^{2n}} = 1 \quad (33)$$

Substituting for C_2 from the definition in Eq. (27),

$$\frac{2an}{a^2\bar{p}_e^{2n}} \frac{RT_f A_t C_D}{VMg} = 1 \quad (34)$$

Letting V/A_t equal L^* and rearranging Eq. (34),

$$L^* = \left(\frac{2aC_D RT_f n}{Ma^2 g} \right) \bar{p}_e^{-2n}$$

$$= \left[\frac{4aC_D RT_f(n/2)}{Ma^2 g} \right] \bar{p}_e^{-2n} \quad (35)$$

For a given propellant burning in its low-pressure region the bracketed term in Eq. (35) can be treated as a constant, and Eq. (35) predicts a straight line correlation of L^* and $\bar{p}_e$ on a log-log plot. Comparing Eqs. (1) and (35), it can be seen that aside from $(\tau_n)_{cr}$ being replaced by $n/2$, they are identical. For JPL-534 propellant, $n/2$ is greater than the value of $(\tau_n)_{cr_1}$ calculated in Ref. 1 by approximately a factor of 2.

6. Low-Pressure Limit for Stable Combustion

Eq. (35) can also be obtained from the non-steady-state ballistic Eq. (28). Setting $r_1 = \bar{r} = a\bar{p}^n$ as previously given, Eq. (28) becomes

$$\frac{dp}{dt} = \frac{C_1 a\bar{p}^n - C_2 \bar{p}}{1 - C_1 \frac{2kn}{c_p \rho a\bar{p}^{n+1}}} \quad (36)^1$$

Again quoting Ref. 5, as $\bar{p}$ becomes small "... the denominator in Eq. (36) becomes small because the negative term in the denominator is an inverse function of pressure, and at the critical pressure $\bar{p}^* = (C_1 \, 2kn/\rho c_p a)^{1/1+n}$ the denominator becomes zero and dp/dt becomes 'infinite'." Near or below $\bar{p}^*$, according to von Elbe (Ref. 5), "the pressure-generating term of the ballistic equation ... tends to 'run away' and more or less violent oscillations result. This instability condition is generally known as 'chuffing' or 'hang-fire' ..."

Setting $dp/dt = 0$ and cancelling like terms in the ballistic Eq. (27) yields the steady-state conditions at equilibrium,

$$A_b r \rho - \frac{C_D \bar{p} A_t}{g} = 0 \quad (37)$$

Again setting $\bar{r} = a\bar{p}^n$,

$$A_b a\bar{p}^n \rho - \frac{C_D \bar{p} A_t}{g} = 0$$

or

$$A_b a \rho - \frac{C_D A_t \bar{p}^{1-n}}{g} = 0 \quad (38)$$

Solving for A_b/A_t,

$$\frac{A_b}{A_t} = \frac{C_D}{a\rho g} \bar{p}^{1-n}$$ (39)

Substituting for C_1 in the expression for the critical pressure $\bar{p}^*$, raising both sides of the expression to the $1 + n$ power, and rearranging terms,

$$\frac{V}{A_b} = \frac{2an\rho RT_f}{Ma} \bar{p}^{*-(1+n)}$$ (40)

Taking the product of Eqs. (39) and (40), where Eq. (39) now expresses the A_b/A_t condition at the critical pressure $\bar{p}^*$,

$$\frac{V}{A_b}\frac{A_b}{A_t} = \left(\frac{2an\rho RT_f}{Ma}\right)\left(\frac{C_D}{a\rho g}\right)\bar{p}^{*-(1+n)} \bar{p}^{*1-n}$$

or

$$\frac{V}{A_t} = L^* = \left[\frac{2aC_D RT_f n}{Ma^2 g}\right]\bar{p}^{*-2n}$$ (41)

It would appear, since Eqs. (35) and (41) are identical, that the critical pressure for solid rocket motor low-pressure combustion extinction is also the theoretical limiting chamber pressure for stable combustion for a given propellant and L^*. Restating the criterion for extinction of combustion, extinction occurs when $f \leq 2\, dq/dt$.

From Eqs. (19), (23), and (25) for a negative dp/dt,

$$2\frac{dq}{dt} = \frac{2k(T_s - T_o)}{r_1^2}\frac{\partial r_1}{\partial p}\left|\frac{dp}{dt}\right|$$

$$= \frac{2kn(T_s - T_o)}{a}\bar{p}^{-(1+n)}\left|\frac{dp}{dt}\right|$$

$$= \infty \text{ at } \bar{p} = \bar{p}^*, \quad \text{since } \left|\frac{dp}{dt}\right| = \infty$$ (42)

At the critical pressure $\bar{p}^*$ the rate of heat flux f to the propellant surface must be infinite to maintain combustion, and therefore the limiting pressure for stable combustion is indeed the pressure at which combustion extinction must occur. This agrees with experimental observations (Ref. 2) which indicated that attempts to ignite motors under atmospheric conditions where the steady-state chamber pressures would be below their extinction pressures resulted in misfires, hangfires, or a chuffing form of burning. There is, therefore, an answer to the

question — why does combustion cease in a nozzled rocket motor when the same propellant will continue to burn at much lower pressures in a strand burner? The answer is: Combustion will cease in a nozzled rocket motor and continue in a strand burner for the same reason that chuffing will occur when an attempt is made to burn the propellant in the motor at chamber pressures below the extinction pressure, whereas no chuffing will occur in the strand burner at these pressures.

7. Estimation of Possible Errors in Analysis

Finally, an estimation has been made of the possible errors in the results of this analysis brought about by some of the simplifying assumptions made. The first assumption was that the heat of gasification Q could be neglected compared to the sensible heat $c_p(T_s - T_o)$ in order to simplify mathematics. For JPL-534 propellant, $Q = 140$ cal/g (Ref. 1) and $c_p(T_s - T_o) = 179$ cal/g from a calculation made using thermophysical properties given in Ref. 1. If the two quantities were equal in magnitude, the bracketed term in the L^*, $\bar{p}_e$ expression, Eq. (35), and the L^*, $\bar{p}^*$ expression, Eq. (41), would each be increased by a factor of 3/2.

In this analysis the erosive component of the total burning rate was neglected and only the pressure dependent linear component considered. In Ref. 6 the following expression for the solid propellant erosive burning rate has been derived:

$$r_e = \gamma G^{0.8}/L^{0.2} \exp\left[\beta\rho(\bar{r}_p + r_e)/G\right]$$ (43)

where γ contains the propellant gaseous heat capacity, viscosity, and Prandtl number and is approximately constant for solid propellants, β is a proportionality constant for a given propellant type, L is the distance along the grain to the point where r_e is being computed, and G is the rocket port mass velocity. From Eq. (43) it is seen that r_e will increase with increasing distance along the grain, due to mass addition increasing the mass velocity, and decrease with increasing linear burning rate.

The small cylindrical propellant grains used in the low-pressure unstable combustion studies (2½-in. diameter × 4- to 4½-in. length) resulted in low port mass velocities and therefore probably low erosive velocity components. A sample calculation of the erosive burning rate component at the nozzle end of the lateral propellant surface just prior to combustion extinction is given for a typical test firing using JPL-534 propellant. The mass velocity at this point is approximated by the calculated total

steady-state mass flow rate divided by the port area, and the erosive burning rate component is neglected in the exponential term of Eq. (42), both approximations increasing the calculated r_e. The modified form of Eq. (43) is

$$r_e = \gamma \left(\frac{\bar{p}_e A_t C_D}{A_p} \right)^{0.8} / L^{0.2} \exp(\beta \bar{r}_p \, \rho A_p / \bar{p}_e \, A_t C_D) \qquad (44)$$

The values used in Eq. (43) are:

$\gamma = 0.24$ in.$^{2.8}$/lb$^{0.8}$ sec$^{0.2}$ (Ref. 6)

$\beta = 10$ (Ref. 6)

$C_D = 6.7 \times 10^{-3}$ 1/sec   (Ref. 1)

$\rho = 0.0625$ lb/in.^{3}

$\bar{r}_p = 0.094$ in./sec   (Ref. 2)

Run 1379 extinction data:

$\bar{p}_e = 32.1$ lb/in.^{2}

$A_t = 0.706$ in.^{2}

$A_p = 4.44$ in.^{2}

$L = 4.16$ in.

The calculated r_e equals 0.0025 in./sec, the total burning rate

$$r = \bar{r}_p + r_e$$

$$= 0.094 \text{ in./sec} + 0.0025 \text{ in./sec}$$

$$= 0.096 \text{ in./sec}$$

and

$$r_e/\bar{r} = 2.6\%$$

It therefore appears that neglecting the erosive burning rate component was a valid assumption for small test motors in a theoretical analysis of low-pressure unstable combustion.

B. Contoured Nozzle Materials and Fabrication Techniques

R. L. Bailey

A contoured nozzle was designed to evaluate the effect on motor performance and compare this performance with previous motor tests that utilized nozzles with an 18-deg exit cone half angle. A submerged nozzle design was used in order to get a direct comparison of similar expansion cone performance.

Based on previous material tests from another program, a carbon cloth–silica cloth phenolic resin combination was used for fabrication. The carbon cloth used was a U.S. Polymeric Chemical Co. material designated as FM-5024. This material had a 35 ± 5% modified phenolic resin system. The silica cloth used was a U.S. Polymeric Chemical Co. material designated FM-5027A and had a 30 ± 3% unmodified phenolic resin system. The materials were applied to a mandrel in a straight, or warped, tape form.

Fig. 2 presents the nozzle design used for fabrication. Since the nozzle was to be tape wrapped to net internal diameter dimensions, a contoured mandrel was required. This mandrel was machined from a steel casting, and then chrome plated to the required internal diameter nozzle dimensions. The mandrel contour was inspected and checked to a plastic template. The nozzle exit diameter was required to have a concentricity of 0.005 in. to the nozzle centerline. The mandrel exit diameter had a concentricity of 0.0005 in., at room temperature, to the nozzle centerline. The mandrel was fabricated so that the nozzle, after wrapping, could be hydroclaved for maximum material density.

Previous tape-wrapped nozzle material tests indicated a tendency for the layers of material to delaminate on cooldown after firing, due to a short bond line (½ in.) between the tape layers. Therefore, a new technique in wrapping was adopted. In order to keep a maximum bond line between the layers of tape, the wrapping angle to the nozzle centerline was varied as the tape progressed from the throat area to the exit area. As shown by Fig. 2, the wrapping angle varied from 0 deg (parallel to the nozzle centerline) in the throat region, to 10 deg through the middle portion of the exit core, and back to 0 deg towards the end of the exit core.

Two nozzles were fabricated by CTL-Division of Studebaker Corporation. These nozzles were hydrotested to 500 psi to check the attachment joint of the steel ring to the nozzle body. The nozzles were also radiographically inspected to insure that there were no voids, porous areas, foreign material, tape delaminations, or wrinkles in the nozzle bodies. The only machining required on the internal surface of the nozzle was in the throat region in order to bond the graphite throat insert. The nozzles were inspected dimensionally; the more pertinent results are listed in Table 2. Fig. 3 presents a view of one of the fabricated nozzles.
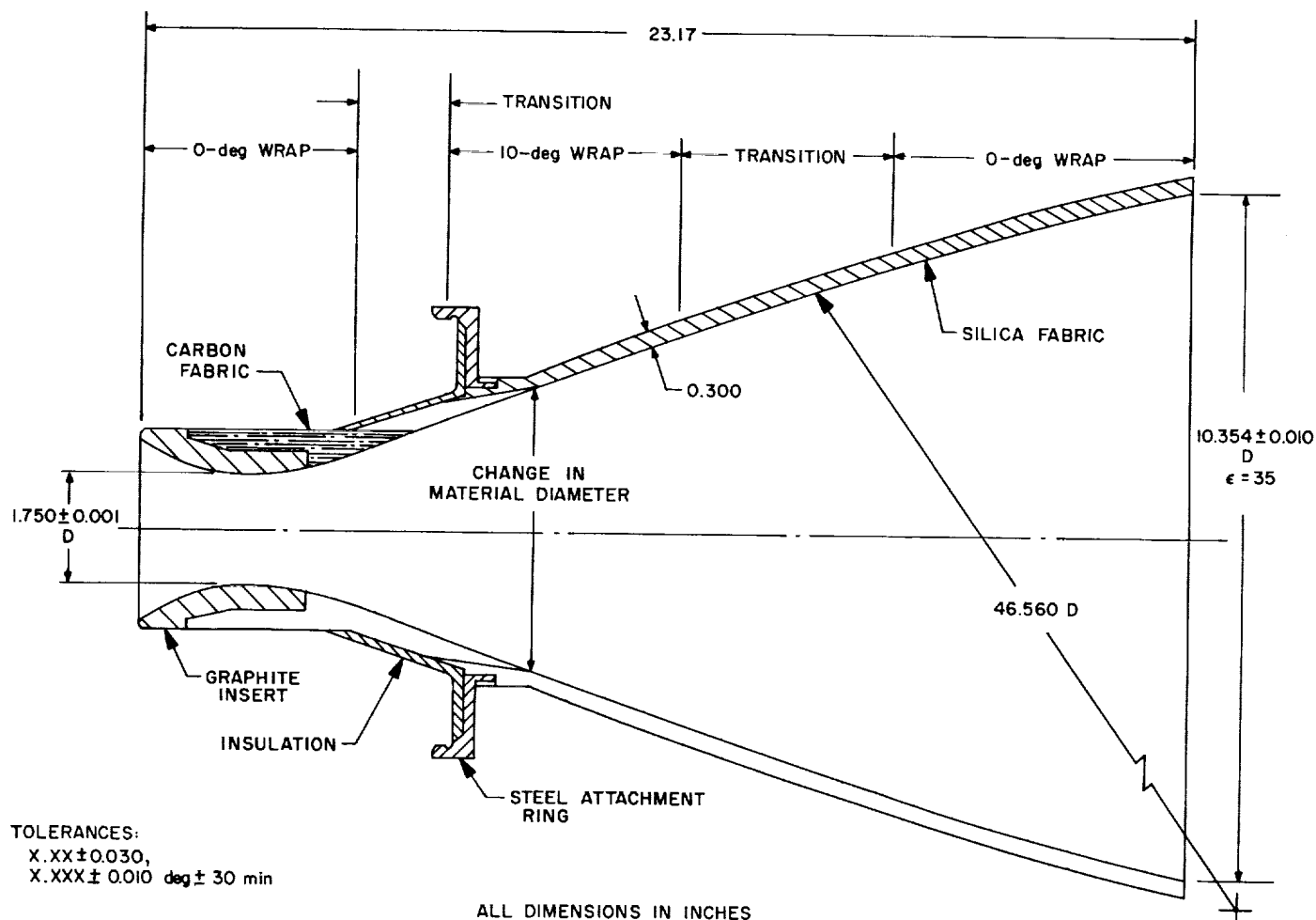
Fig. 2. Contoured nozzle design

The motors utilized for the tests were the same as those used to test 18-deg exit cone half angle nozzles. These motors had a polyurethane-ammonium perchlorate 16% aluminum solid propellant, and provided a 20-sec burning time. Both motors were fired successfully in the altitude facility at the JPL Edwards Test Station. Fig. 4 shows a typical test setup in the altitude test chamber.

The motor performance agreed favorably with previous motor tests of 18-deg nozzle cone configurations, which had the same throat diameter, exit diameter ($\epsilon = 35$), and were tested on similar motors in the same altitude facility. The contoured nozzle performance, as expected, agreed within 0.5% of the 18-deg conical nozzle performance.

Table 2. Contoured nozzle dimensions

Region	Required dimensions, in.	Actual dimensions, in.	
		SN-1	SN-2
Throat diameter	1.750 ± 0.001	1.7510	1.7501
Throat concentricity	0.005 to nozzle CL	0.00013	0.00010
Exit contour radius	46.560 ± 0.010	Checked to template	Checked to template
Exit diameter	10.354 ± 0.010	10.355	10.357
Exit concentricity	0.005 to nozzle CL	0.0035	0.0020
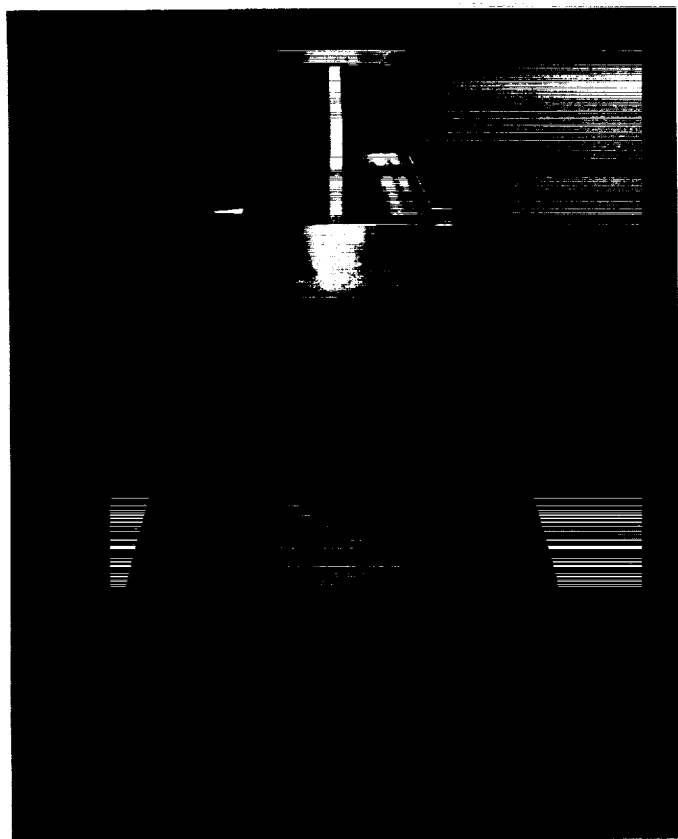
Fig. 3. Contoured nozzle after fabrication

The nozzles performed very well in that there was no delamination of tape, negligible cone erosion, and no gouging or tearing of material. The nozzles were inspected after firing; the results are shown in Table 3.

Table 3. Contour nozzle firing results

Region	Dimension before firing, in.		Dimension after firing, in.	
	SN-1	SN-2	SN-1	SN-2
Throat diameter	1.7510	1.7501	1.766	1.7555
Throat concentricity	0.00013	0.00010	0.004	0.001
Exit diameter	10.355	10.357	10.311	10.313
Exit concentricity	0.0035	0.0020	0.011	0.006

The nozzles were sectioned in order to inspect the charring pattern. The char depth was very similar to 18-deg cone firings of similar materials. It was very uniform with no excessive charring occurring in any one location due to the contour. Fig. 5 shows one of the sectioned nozzles.

Fig. 4. Contoured nozzle test setup
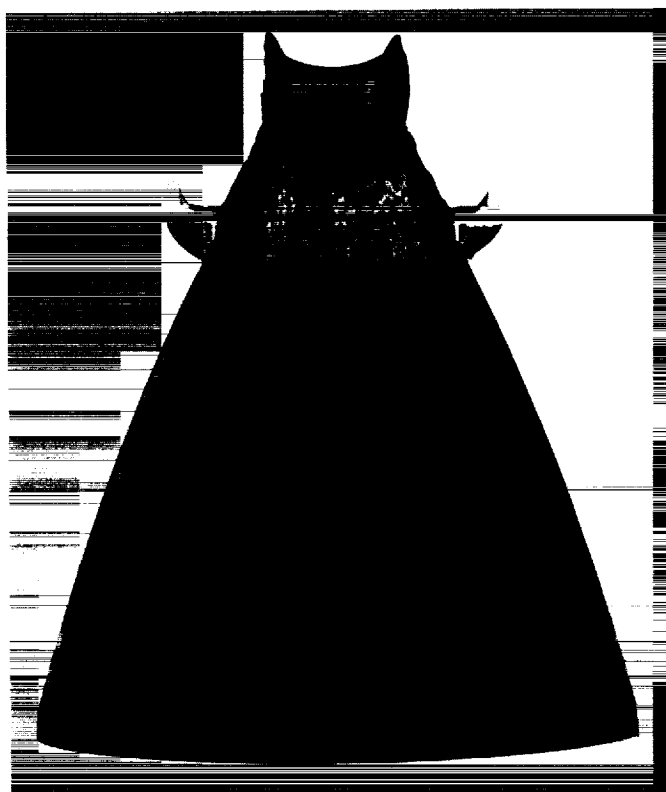
Fig. 5. Sectioned contoured nozzle showing charring pattern

C. Nozzle Materials Evaluation Program

R. L. Bailey

A materials evaluation program was initiated to test various materials suitable for use with solid propellant rocket motors. This is a continuing program that has the following objectives:

(1) The evaluation of materials for nozzle and motor insulation applications.

(2) The development of material fabrication and quality control procedures.

(3) To provide support for research and motor development programs.

The materials were tested on motors that were loaded with a polyurethane-ammonium perchlorate propellant with 16% aluminum giving a flame temperature of approximately 6000°R. The motor burning time was, or could be, varied from 1 to 75 sec.
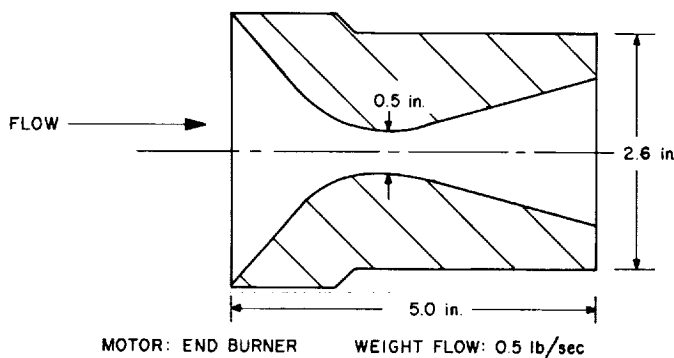
Fig. 6. Molded nozzle for material evaluation
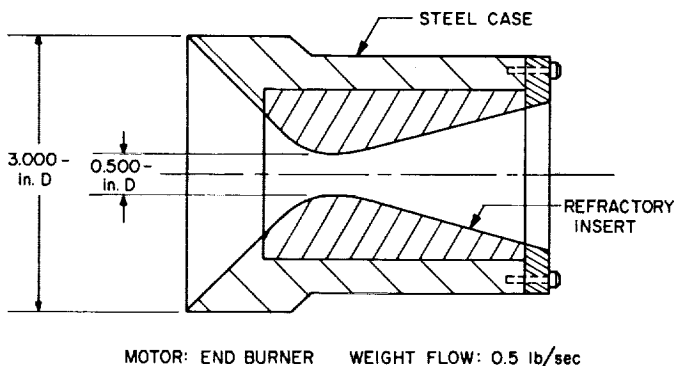
Fig. 7. Nozzle configuration for testing refractory throat inserts

Table 4. Typical plastic material composition data

Type[a]	Phenolic resin weight, 1 %	Cloth material[b]	Specific gravity
FM-5049	50	Carbon cloth	1.44
MX-3555	27–31	wheat grains Chopped silica	1.89
FM-5020	28	Chopped silica	1.70
MX-4551	40	Chopped graphite	1.42
FM-5014	34	Chopped graphite	1.42
SL-401	50	Macerated carbon	1.40.

[a]FM: U.S. Polymeric Chemicals, Inc.
MX: Fiberite Corporation
SL: Thermal Materials, Inc.
[b]Chopped: cloth is cut into ½ × ½-in. squares
Wheat grain: cloth is cut into small particles which resemble grains of wheat
Macerated: cloth is cut into many short fiber lengths, ½ to ¾-in. long

The type of materials evaluated were plastics, metals, ceramics, refractories, and elastomerics. These materials were tested in nozzle configurations shown in Figs. 2[1], 6, and 7. Table 4 presents the results of the materials tested to date.

1. Plastic Materials

The plastic materials were first tested in the configuration shown in Fig. 6. Then the more successful ones were tested in the configuration shown in Fig. 7, or Fig. 2[1]. The materials were molded in (1) fiber-oriented configurations, or (2) random fiber orientation as obtained with macerated, chopped, or bulk fiber form; then they were tested. Table 4 shows the composition of some typical plastic materials used in the evaluation. As shown by Table 5 the carbon-cloth material exhibited the lowest throat erosion rate, while the silica-cloth material exhibited the lowest char rate. The lowest erosion rate was obtained with carbon cloth molded in the macerated form.

In an attempt to show the variation of erosion rate with burning time, four macerated carbon-cloth nozzles were molded to the Fig. 6 configuration. These nozzles were tested with the same propellant and chamber pressure but the burning times were varied from 1 to 27 sec. The data obtained (Fig. 8) indicated that the nozzle throat contracted for a period of time before any erosion occurred.

2. Refractory Materials

The refractory materials were tested as nozzle throats in the configurations shown by Figs. 2 and 7. Of all the

---

[1] See p. 94, preceding article.

Table 5. Results for nozzles tested with aluminized propellant

Material[a,b]	Initial throat diameter, in.	Initial pressure, psia	Burning time, sec	Throat[c] erosion rate, mils/sec	Typical char rate, mils/sec
Molded Plastic					
Silica					
MX-3555	0.500	230	22	11.76	2.33
MX-1344-67	1.500	500	5.3	29.00	8.33
REPCO-200	0.499	230	28	8.32	3.33
FM-5048	0.499	230	34	8.80	5.14
MX-3581	0.500	200	30	10.13	—
AT-1	0.498	200	24	10.15	3.84
FM-5020	0.499	200	29	8.35	—
Graphite					
MX-4551	0.500	230	17	14.20	11.11
FM-5014	0.499	200	28	1.89	8.51
REPCO-400	0.499	230	20	2.80	—
Carbon					
FM-5049	0.499	200	25	2.48	13.71
SL-401	0.500	200	45	0.30	7.11
SL-406	0.500	200	29	0.56	—
SL-407	0.499	200	29	0.11	—
SL-437	0.500	200	26	0.08	9.00
SL-401 Tape 90°—C.L.	0.500	200	47	0.30	—
SL-401 Tape 90°—C.L.	1.082	1000	6.3	2.0	14.28
SL-431	1.382	500	6	4.0	—
SL-401	1.750	250 max.	19.5	1.0	7.61
Solid Graphite					
Pyro-graphite (H.T.M.)	0.349	400	37	0.028	—
ZTA	1.750	250 max.	19.5	0.09	—
Graph-i-tite "G"	1.750	250 max.	19.5	0.06	—
HLM-85	0.435	300	27	<0.001	—
Refractory					
Ti-BN	0.524	200	46	<0.001	—

[a]All the plastic materials tested utilized a phenolic resin system.
[b]REPCO: Reinhold Engineering and Plastic Company
MX: Fiberite Corporation
FM: U.S. Polymeric Chemicals, Inc.
AT-1: Aerothermal, Inc.
SL: Thermal Materials, Inc.
HLM: Great Lakes Carbon Company
H.T.M.: High Temperature Materials
ZTA: National Carbon Company
"G": Graphite Specialties Corporation
[c]Erosion rate is based on the throat diameter.

throat materials tested the pyro-graphite exhibited the lowest erosion rate. A Ti-BN (titanium-boron-nitride) system, although it had a thick coating (0.030 in.) after firing, showed excellent capability as a heat sink material, and exhibited negligible erosion. All the bulk graphite tested had a comparable density of approximately 1.88 g/cm³.
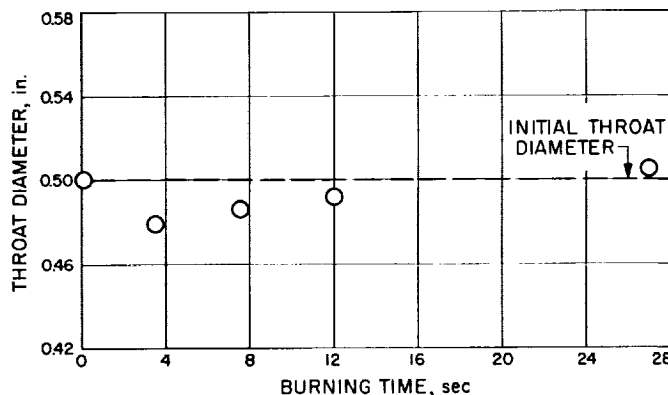
Fig. 8. Variation of nozzle throat diameter with burning time

D. Sterilization of Solid Propellants

L. C. Montgomery

As indicated in Ref. 7 a chemical system for sterilizing a rocket motor has been developed. However, chemical sterilization methods have not progressed to the point of being acceptable for sterilizing the entire spacecraft. Therefore, JPL has turned to heat sterilization as a means of meeting sterility standards. To meet the requirements of heat sterilization an investigation of "off the shelf" propellant systems has been started.

The initial steps in this investigation began with inquiries to all solid propellant manufacturers in the United States indicating the criteria for heat sterilization and the other desired characteristics for such a propellant. These are given here in descending order of importance:

(1) The propellant must have high temperature stability and retain acceptable physical and ballistic properties and have high reliability after three heat sterilization cycles of 145°C for 36 hr in a sealed container.

(2) The propellant must be capable of being ignited in a vacuum after sterilization and six months of storage of the propellant.

(3) The propellant must retain its physical and ballistic characteristics for six months storage after sterilization.

(4) The exhaust products must be as free as possible from solid products.

(5) Physical properties must be such as to allow manufacture of a motor having high reliability.

(6) The specific impulse of the propellant is not a chief concern since the present investigation is concerned with propulsion units on the order of 5 lb.

In response to JPL inquiries, request for proposals, purchase orders, etc., six companies and one government agency have indicated that one of their propellants should satisfy the above requirements. Four of these propellants have been subjected to the initial phase of screening tests, two more have recently been received, and one not yet supplied.

The initial test performed or to be performed on the candidate propellants is the exposing of the propellant to the three sterilization temperature cycles and observing physical changes that occur. This was done by placing samples $\frac{1}{2} \times \frac{1}{2} \times 4$ in. long cantilevered from one end, as shown in Fig. 9, into small "disposable" ovens constructed by JPL for this purpose. Then records and photographs are made to indicate physical changes that occur.

Results of the first series of tests are shown in Figs. 9a through 9d. All of these propellants, except that shown in Fig. 9d, have undergone the full three cycles of 36 hr at 145°C. Initial indications are that one propellant is unchanged by sterilization; two propellants show some change in physical appearance but may still be acceptable in particular motor configurations. The fourth propellant melted in the oven during the first cycle.
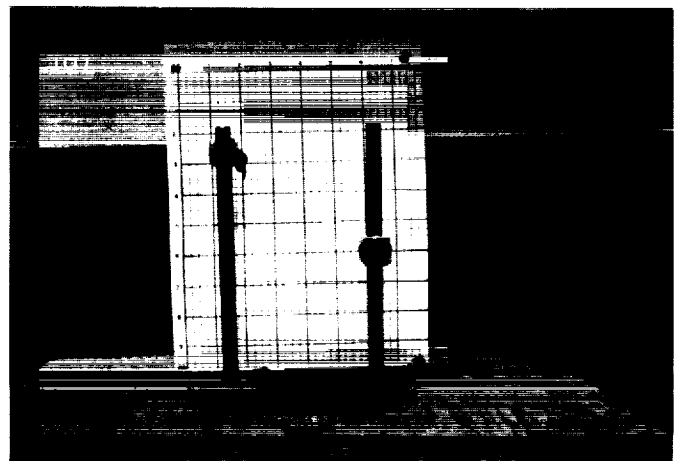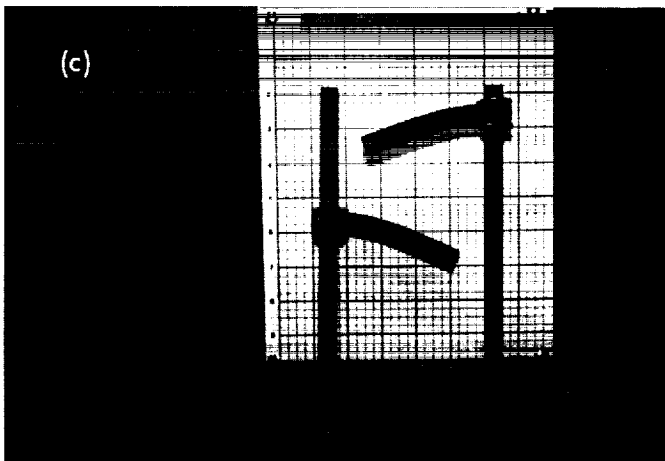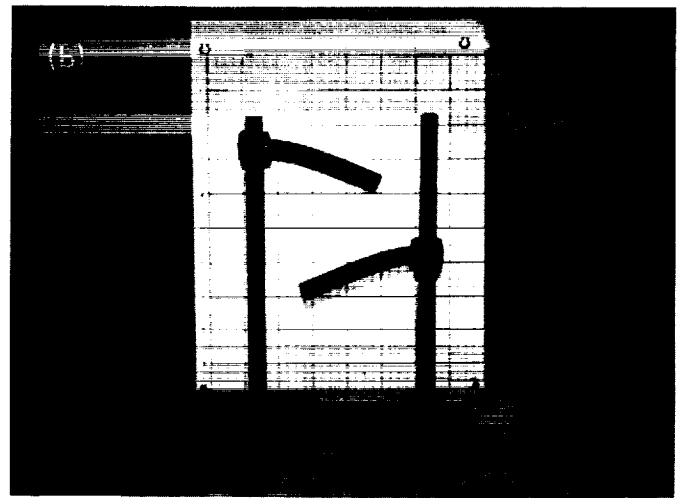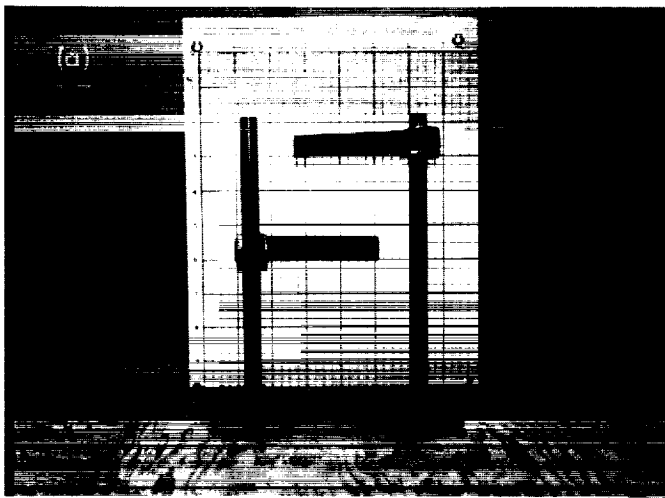
Fig. 9. Results of heat sterilization tests on various solid propellants

Screening tests will be continued on the surviving propellants and those just received.

Further screening tests will consist of determining the effect of sterilizing in inert atmospheres, sterilizing of physical test specimens with different amounts of exposed surfaces, and physical dimension changes in a large block of propellants.

For the propellant candidates that survive the screening tests further evaluation will be made through more precise physical tests, ballistic tests, ignition tests, tests of space environment effects, tests of the effects of exposure to chemical sterilization gases, and storage and aging tests.

E. One-Dimensional Analysis of Gas-Particle Flow in Solid Propellant Rocket Motors

C. F. Robillard

1. Introduction

The traditional thermodynamic rocket performance calculation programs assume isentropic expansion of the products of combustion. If the combustion products contain a condensed phase, such as Al_2O_3, the assumption of an isentropic expansion requires, among other things, that the condensed particles be in velocity and temperature equilibrium with the gas.

For propellants which yield condensed phases in the combustion products, a more realistic performance calculation requires a step-wise numerical solution of the differential equations governing gas velocity, particle velocity, gas temperature, and particle temperature along the nozzle. The first such set of general, one-dimensional gas-particle flow equations was published by Kliegel (Ref. 8). An IBM 7094 program based on these equations has been developed at JPL to provide a fluid dynamic estimate of propellant performance as a function of:

(1) Weight fraction of condensed phase in the exhaust, α.

(2) Particle radius r_p and density m_p.

(3) Nozzle design and expansion ratio:

    (a) Entrance angle θ and exit angle ϕ.

    (b) Nozzle throat diameter D_t and radius of curvature R_t.

(4) Chamber pressure P_c.

(5) Gas thermodynamic properties.

These parameters were systematically varied, as indicated in Table 6, to examine their effect on performance efficiency.

2. Computer Program

A complete description of the fluid dynamic (See Footnote 1, following page) performance calculation program will be given in a separate report. The most important assumptions are the following conventions:

(1) The particles occupy negligible volume, and exert no pressure.

(2) The chemical composition of the gas is frozen.

Table 6. Range of variables

Composition and aluminum content			
Propellant designation	Wt % ammonium perchlorate	Wt % binder	Wt % Al
AI5AC	76	19	5
AI5AA	69	19	12
AI5AD	61	19	20

Chamber pressure: 100, 500, 1000 psi.
Radius of aluminum oxide particles: 0.5, 1.0, 1.5, 2.0, 3.0, and 4.0 μ.
Density of aluminum oxide particles: 3.5 and 4 g/cm³.

Nozzle geometry (maximum ε ≅ 100)				
Nozzle designation	Entrance angle θ, deg	Exit angle φ, deg	Throat diameter D_t, in.	Throat radius of curvature R_t, in.
11	30	15	0.5	0.5
12	30	15	0.5	1.5
15	30	30	0.5	0.5
1	30	15	1.24	1.24
13	30	15	6.0	6.0
16	30	15	6.0	3.0
18	80	30	6.0	3.0
19	30	15	1.24	9.92

Only conical nozzles with constant radii of curvature were considered.

(3) Condensation of additional particles during expansion is not permitted. Therefore, the weight fraction of particles, α, is constant along the nozzle.

(4) Particle solidification is permitted.

(5) The particles are spherical and of uniform temperature.

(6) The particles transfer energy to the gas through convection and this is the only mechanism by which the particle temperature decreases.

(7) The particle is accelerated as a result of drag forces between the particle and the moving gas.

The first four assumptions above are also embodied in the thermodynamic[1] program calculation of frozen flow performance.

3. Qualitative Results from the Fluid Dynamic Performance Calculation Program

a. Summary. The ratio of the fluid dynamic[1] c^* or I_{sp} to its thermodynamic[1] counterpart c^{*0} or I_{sp}^0 is a measure of the calculated performance loss due to velocity and thermal lag of condensed particles in the gas. This ratio is termed the two-phase-flow efficiency.

Qualitatively, the I_{vac} efficiency at any epsilon ϵ and the c^* efficiency decrease as:

(1) The weight fraction of Al_2O_3, α, increases.

(2) The particle radius r_p increases.

(3) The throat diameter D_t decreases.

(4) The chamber pressure P_c decreases.

(5) The exit angle ϕ increases (I_{vac} only).

(6) The particle density m_p increases.

(7) The nozzle radius of curvature at the throat R_t decreases.

(8) The entrance angle θ increases.

I_{vac} efficiency may increase and/or decrease with increasing epsilon ϵ.

The above parameters have been listed in approximately decreasing order of maximum effect on efficiency, based on the range of values considered, e.g., the maximum change in efficiency due to varying particle radii from 0 to 4 μ was greater than the maximum change in efficiency due to varying the entrance angle from 30 to 80 deg, the chamber pressure from 1000 to 100 psi, or the particle density from 3.5 to 4.0 g/cm³. Actually, the magnitude of the efficiency decrease associated with varying any one parameter is very much a function of the other parameter values, and the magnitude increases as the other parameter values become less favorable. For example, the decrease in efficiency due to increasing the particle radius from 1 to 2 μ is larger at $D_t = 0.5$ in. than at $D_t = 4$ in., or larger at $P_c = 100$ psi than at $P_c = 1000$ psi.

Many of the calculations have been performed using two drag laws, denoted by A and B. Qualitatively, the results are the same. Quantitatively, Drag Law B is generally less favorable. Using the results of the IBM 7094 calculations for Drag Law A, a simplified set of empirical equations have been obtained which reproduce c^* and I_{vac} ($\epsilon = 1$ to 40) values from the 7094 calculations with a maximum error of 1% in I_{vac} or c^* and a mean error of about 0.5%. The next few sections discuss observations which led to combining or eliminating some variables and hence to the empirical equations.

b. The effect of wt% Al_2O_3 (α). For a homologous series of propellants[2], the two-phase flow efficiency (*Eff*) was found to decrease almost linearly with increasing α, that is:

$$Eff = 1 - k\alpha$$

This relationship had been discovered empirically by Ingham (Ref. 9) from a perusal of motor firing data.

For a homologous series of propellants, the slope k is not independent of α; however, it is remarkably insensitive, as shown in Figs. 10 and 11. The discrepancies at $\epsilon = 20$ are apparently associated with particle freezing

---

[1] Fluid dynamic refers to the step-wise numerical solution of the one-dimensional two-phase flow equations. Thermodynamic refers to the path-independent solution obtained by assuming the expansion to be isentropic.

[2] For this particular series of propellants, the binder content was held constant and aluminum was substituted for ammonium perchlorate. The basic criteria for an homologous series of propellants is discussed later.
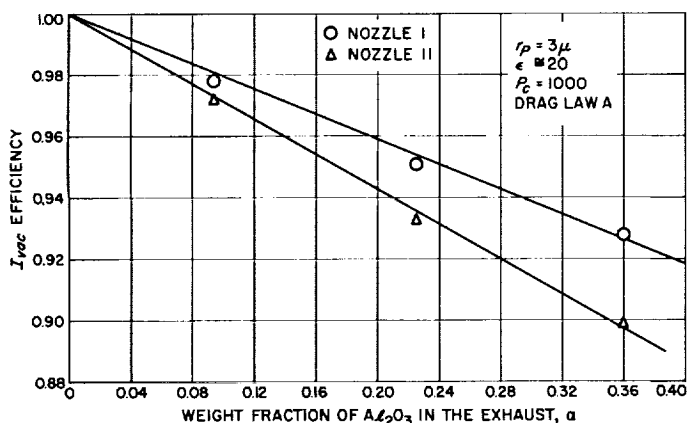
Fig. 10. I_{vac} **efficiency vs** α
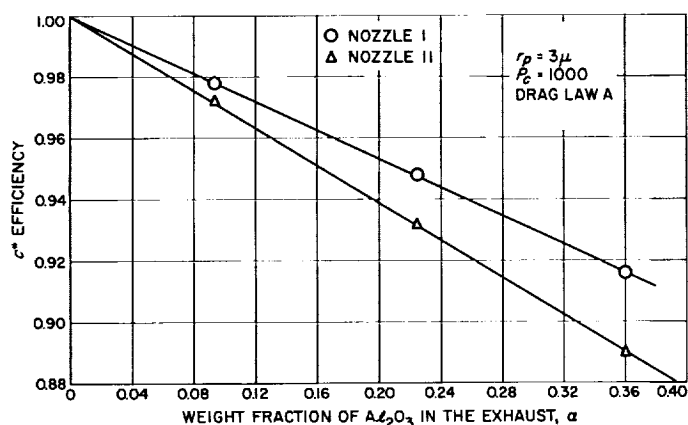
Fig. 11. c^* **efficiency vs** α

and they dampen out further down the nozzle. Note that the relationship is most accurate when the slope k used is that of the maximum value of α; that is, for a given error in k, the corresponding error in efficiency decreases as α decreases. The obvious advantage of the relationship is that the analytical study can be confined to one value (the maximum value) of α.

The slope k is then a function of:

(1) Particle radius r_p and density m_p.

(2) Nozzle geometry and epsilon ϵ.

(3) Chamber pressure P_c.

(4) Gas thermodynamic properties.

Note also that the value of k for c^* Eff is not the same as the corresponding value of k for I_{vac} Eff at $\epsilon = 1$.

While c^* is dependent only on the fluid properties at the throat, I_{vac} represents the integrated properties all along the nozzle path. The c^* Eff is lower than the I_{vac} Eff at $\epsilon = 1$ because the velocity lag is continually increasing $(V_p/V_g$ is decreasing) from the nozzle entrance to the throat.

c. The influence of the gas thermodynamic properties.
The limiting values of k are:

$$k = 0 \quad \text{for } r_p = 0 \quad (\text{or } r_p > 0, D_t \to \infty)$$

$$k = A \text{ for } r_p \to \infty \quad (\text{or } r_p > 0, D_t = 0)$$

so that

$$c^* \, Eff_{(r_p \to \infty)} = 1 - A\alpha = \frac{c^*_{(r_p \to \infty)}}{c^{*0}}$$

and

$$I_{vac} \, Eff_{(r_p \to \infty)} = 1 - A\alpha = \frac{I_{sp\,(r_p \to \infty)}}{I^0_{sp}}$$

Conceptually, I_{vac} and $c^* \, Eff_{(r_p \to \infty)}$ correspond to the case where all of the condensed particles are left behind in the motor chamber. Therefore, A is a function of the gas thermodynamic properties alone, and it varies only with chamber pressure and epsilon for any given propellant.

For the homologous series of propellants investigated, the slope A showed the same lack of sensitivity to α as did k. This observation, together with the confirming observation that, for plots of k versus r_p or k versus D_t, k approached the same asymptotic value (namely A), indicated that close-to-equal values of A are the basic criteria for an homologous series of propellants. In other words, the two-phase flow efficiency decreases almost linearly with increasing α for a series of propellants exhibiting close-to-equal values of A.

Because k increases and efficiency decreases with increasing A, propellants with low values of A (high performance of the gas at $r_p \to \infty$) should exhibit higher efficiencies, all other parameters being the same.

d. The effect of Al_2O_3 particle radius and density.
Particle size analysis of Al_2O_3 collected from motor firings show particle-size distributions which appear to vary with motor firing conditions and motor size. The calculation of effective particle size from a particle-size distribution is

discussed later. The range of particle sizes considered here includes all the effective particle sizes reported to date but not the maximum size particles in all distributions.
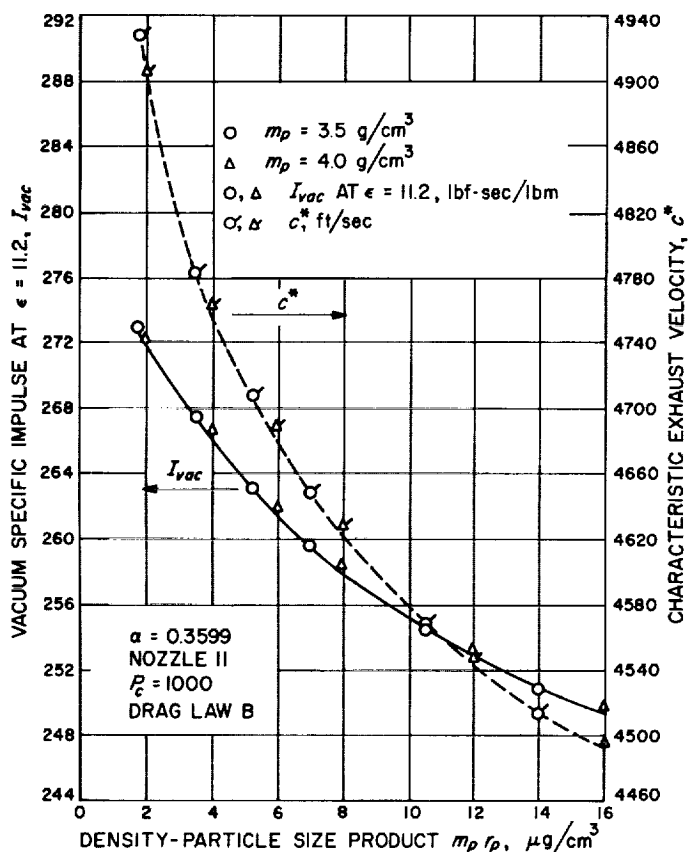
Even for Al_2O_3 particles of uniform cross section (no shells), there is some uncertainty in the density of both liquid and solid phases at high temperatures. This uncertainty is reflected in the range of densities considered.

Sample plots of I_{vac} and c^* as a function of the $m_p r_p$ product are shown in Fig. 12. Within the range of particle radii and densities considered, percentage variations in m_p and r_p are almost interchangeable; that is, for maximum errors in efficiency of about 0.3%, I_{vac} and c^* are constant for values of the $m_p r_p$ product. More precisely, however, a given percentage increase in m_p lowers I_{vac} and c^* less than does the same percentage increase in m_p.

e. The effect of nozzle geometry.

Geometry of the nozzles employed. Only conical, converging-diverging nozzles with throats of constant radius of curvature were considered in this study. Their geometry is shown in Fig. 13. The initial radius (r_0) was constrained to a value of 1.9 D_t.

Effect of entrance angle. Only one large entrance angle (80 deg for Nozzle 18) was used in this series of calculations. Calculations performed prior to this series indicated that the effects on I_{vac} and c^* of varying the entrance angle from 30 to 80 deg were small (less than 0.5%) and nearly independent of throat diameter. The small change in performance for the large change in entrance angle can be explained by observing that the foreshortening occurred in the nozzle region of low acceleration.

Fig. 12. I_{vac} **and** c^* **as a function of the** m_p r_p **product**
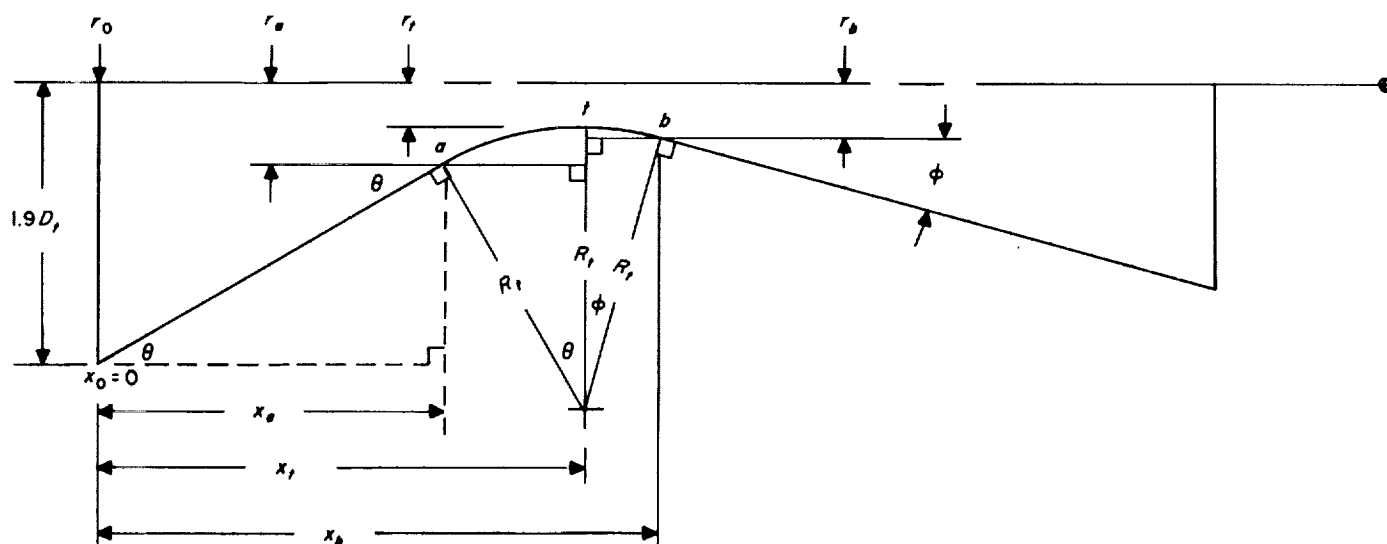
Fig. 13. Nozzle geometry

A nozzle geometry function for c*. It was found that the total effect of nozzle geometry on c^* *Eff* could be combined in a single nozzle geometry function, to wit, the epsilon value ϵ_x, at an arbitrary and equal Δ_x from the throat. Following is the derivation of ϵ_x as a function of R_t, D_t, and Δ_x:
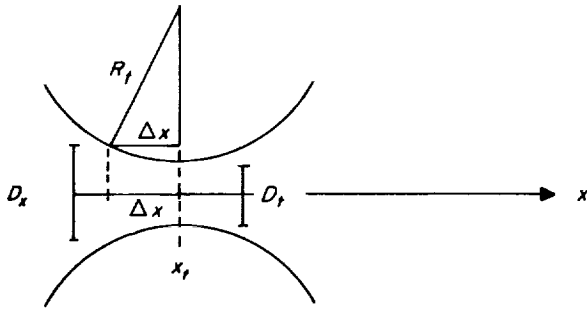
$$\epsilon_x = \frac{D_x^2}{D_t^2}$$

$$\frac{D_x}{2} = (\epsilon_x)^{\frac{1}{2}} \frac{D_t}{2} = r_x$$

$$(\epsilon_x)^{\frac{1}{2}} \frac{D_t}{2} + (R_t^2 - \Delta x^2)^{\frac{1}{2}} = R_t + \frac{D_t}{2} = R_t + r_t$$

$$(\epsilon_x)^{\frac{1}{2}} = \frac{R_t + \frac{D_t}{2} - (R_t^2 - \Delta x^2)^{\frac{1}{2}}}{\frac{D_t}{2}}$$

$$= \frac{2R_t}{D_t} + 1 - \frac{2}{D_t}(R_t^2 - \Delta x^2)^{\frac{1}{2}}$$

$$\frac{(\epsilon_x)^{\frac{1}{2}} - 1}{2} = \frac{R_t}{D_t}\left[1 - \left(1 - \frac{\Delta x^2}{R_t^2}\right)^{\frac{1}{2}} \right]$$

Note that the nozzle geometry function has been defined as $[(\epsilon_x)^{\frac{1}{2}} - 1]/2$ rather than ϵ_x itself. The restriction on Δx is that Δx be less than the minimum value of R_t. For these calculations, Δx was taken to be 0.4 in.

Fig. 14 shows the nozzle geometry function versus a slope function $(Ak/A - k)$ where

$$c^* \; Eff = 1 - k\alpha$$

and A is the slope k at $r_p \to \infty$. For the three propellants shown in Table 6, the value of the slope A for c^* *Eff* is 0.609. Also shown on the right-hand side of Fig. 14 is the c^* *Eff* scale corresponding to the slope function scale for $A = 0.609$. For only a small error in c^* *Eff*, the slopes of the geometry function curves can be considered independent of particle radius. However, the slopes and intercepts of these curves are pressure dependent.

4. Calculation of Effective Particle Size from a Particle-Size Distribution

Particle-size analysis of Al_2O_3 collected from motor firings shows particle-size distributions which appear to vary with motor firing conditions and motor size. It is possible to perform fluid dynamic performance calculations for a given distribution of particle radii rather than a single particle radius. However, there is a very large number of possible distributions involving a comparatively small number of particle radii, and it would be preferable to perform the calculations for a range of single particle radii and then to calculate an effective average particle radius for each distribution.

The derivation of the equation for calculating the effective average particle radius from a particle size distribution follows:

$$(M_g^2 - 1)\frac{dV_g}{V_g}$$

$$= \frac{dA}{A} - \frac{\dot{w}_p \, MW_g \, V_g}{\dot{w}_g \, g_c \, R \, T_g} \frac{3}{8} \frac{\rho_g \, dx}{m_p} \sum \frac{C_{Di} \, (V_g - V_{pi})^2 \, x_{pi}}{V_{pi} \, r_{pi}}$$

$$+ \frac{\dot{w}_p \, MW_g}{\dot{w}_g \, g_c \, J \, C_g \, T_g} \frac{3}{8} \frac{\rho_g \, dx}{m_p} \sum \frac{C_{Di} \, (V_g - V_{pi})^2 \, x_{pi}}{r_{pi}}$$

$$- \frac{\dot{w}_p \, MW_g}{\dot{w}_g \, C_g \, T_g} \frac{3 \, dx}{m_p} \sum \frac{h_{pi} \, (T_{pi} - T_g) \, x_{pi}}{r_{pi} \, V_{pi}} \qquad (1)$$

$$\sum A = \sum \frac{x_{pi} \, (V_g - V_{pi})^2 \, C_{Di}}{V_{pi} \, r_{pi}} = \frac{(V_g - \bar{V}_p)^2 \, \bar{C}_D}{\bar{V}_p \, \bar{r}_p} \qquad (2)$$

$$\sum B = \sum \frac{x_{pi} \, (V_g - V_{pi})^2 \, C_{Di}}{r_{pi}} = \frac{(V_g - \bar{V}_p)^2 \, \bar{C}_D}{\bar{r}_p} \qquad (3)$$

Let

$$C_{Di} = \frac{12 \, \mu_g}{\rho_g \, (V_g - V_{pi}) \, r_{pi}} ; \quad \bar{C}_D = \frac{12 \, \mu_g}{\rho_g \, (V_g - \bar{V}_p) \, \bar{r}_p} \qquad (4)$$
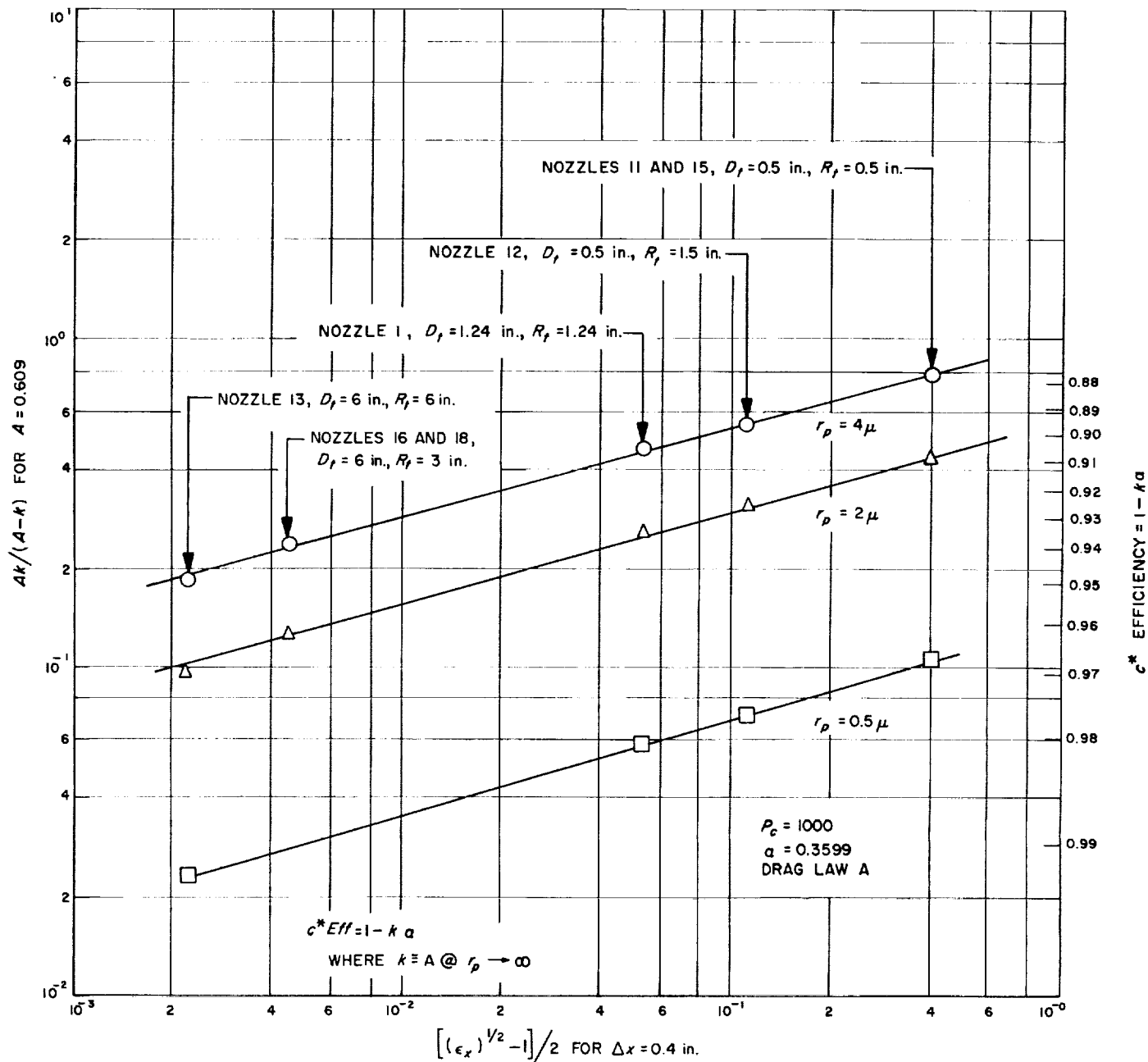
Fig. 14. A nozzle geometry function for c*

Then

$$\sum A = \frac{(V_g - \bar{V}_p)\,12\,\mu_g}{\bar{V}_p\,\rho_g\,\bar{r}_p^2} = \frac{12\,\mu_g}{\rho_g}\sum \frac{x_{pi}\,(V_g - V_{pi})}{V_{pi}\,r_{pi}^2} \qquad (5)$$

$$\sum B = \frac{(V_g - \bar{V}_p)\,12\,\mu_g}{\rho_g\,\bar{r}_p^2} = \frac{12\,\mu_g}{\rho_g}\sum \frac{x_{pi}\,(V_g - V_{pi})}{r_{pi}^2} \qquad (6)$$

From Eqs. 2 and 3,

$$\frac{\Sigma B}{\Sigma A} = \bar{V}_p \qquad (7)$$

Rearranging Eq. 2,

$$\bar{r}_p = \frac{(V_g - \bar{V}_p)^2\,\bar{C}_D}{\bar{\bar{V}}_p\,\Sigma A} \qquad (8)$$

From Eqs. 4, 7, and 8,

$$\bar{r}_p^2 = \frac{\left(V_g - \dfrac{\Sigma B}{\Sigma A}\right)12\,\mu_g}{\dfrac{\Sigma B}{\Sigma A}\,\Sigma A\,\rho_g} \qquad (9)$$

From Eqs. 5, 6, and 9,

$$\bar{r}_p = \left(\frac{1 - \dfrac{\displaystyle\sum \dfrac{x_{pi}\,\Delta V}{r_{pi}^2\,V_g}}{\displaystyle\sum \dfrac{x_{pi}\,\Delta V}{V_{pi}\,r_{pi}^2}}}{\displaystyle\sum \dfrac{x_{pi}\,\Delta V}{r_{pi}^2\,V_g}}\right)^{\!\!\frac{1}{2}} \qquad (10)$$

where $\Delta V = V_g - V_{pi}$.

Table 7. Calculation of effective particle radius (Nozzle 1, $P_c = 1000$, $\epsilon = 9.79$)

				Effective average particle radius from Eq. (10)					
x_{pi}	$r_{pi},$ μ	r_{pi}^2	V_g	V_{pi}	$\Delta V = V_g - V_{pi}$	$x_{pi}\,\Delta V$	$\dfrac{x_{pi}\,\Delta V}{r_{pi}^2}$	$Y = x_{pi}\,\Delta V/r_{pi}^2\,V_g$	$Z = x_{pi}\,\Delta V/r_{pi}^2\,V_{pi}$
0.2	0.5	0.25	8241	8087	154	30.8	123.2	14.95×10^{-3}	15.23×10^{-3}
0.3	1.0	1.0	8109	7701	408	122.4	122.4	15.09	15.89
0.2	1.5	2.25	8169	7401	761	152.2	67.64	8.280	9.139
0.1	2.0	4.0	8233	7082	1151	115.1	28.77	3.494	4.062
0.1	3.0	9.0	8352	6527	1825	182.5	20.28	2.428	3.107
0.1	4.0	16.0	8422	6130	2292	229.2	14.32	1.700	2.336
								$\Sigma Y = 45.94 \times 10^{-3}$	$\Sigma Z = 49.76 \times 10^{-3}$

$$\bar{r}_p^2 = \frac{1 - (\Sigma Y/\Sigma Z)}{\Sigma Y} = \frac{1 - (45.94/49.76)}{45.94 \times 10^{-3}} = \frac{0.0768}{45.94 \times 10^{-3}} = 1.67$$

From Eq. (10), $\bar{r}_p = 1.30$.

Particle radius corresponding to the wt average I_{sp}			
r_{pi}	x_{pi}	I_{vac}	$x_{pi}\,I_{vac}$
0.50	0.2	277.7	55.54
1.0	0.3	271.5	86.45
1.5	0.2	267.7	53.54
2.0	0.1	264.5	26.45
3.0	0.1	260.0	26.00
4.0	0.1	256.5	25.65
		$\Sigma\,x_{pi}\,I_{vac} = 268.63$ = wt average I_{sp}.	

From wt average I_{sp}, $\bar{r}_p = 1.38$.

Weight average particle radius		
x_{pi}	$r_{pi},$ μ	$x_{pi}\,r_{pi}$
0.2	0.5	0.10
0.3	1.0	0.30
0.2	1.5	0.30
0.1	2.0	0.20
0.1	3.0	0.30
0.1	4.0	0.40
	$\Sigma x_{pi}\,r_{pi} = \bar{r}_p = 1.60$	

Eq. (1) is used in the program for calculating the gas velocity. This equation was used as a starting point because it contains all the terms used in the program which involve particle radius. Eq. (4), the simplified equation for drag coefficient, is a fair approximation of the drag law for $P_c = 1000$ used in the program at the time these calculations were made. Eq. (10) indicates that the effective particle radius, $\bar{r}_p$, may well be a function of all the parameters which affect I_{sp} efficiency; that is, for a given particle size distribution, there is no single value of effective average particle radius, but many values.

A sample calculation of effective average particle radius from Eq. (10) is shown in Table 7. Also shown

in Table 7 are the weight-average particle radius and the particle radius which corresponds to the weight average I_{sp}. These three particle radii are compared again in Table 8 for several particle-size distributions. The weight-average particle radius is always significantly larger than the other two.

The effective average particle radius from Eq. (10) and the particle radius corresponding to the weight average I_{sp} are close enough to suggest that the I_{sp} for any particle size distribution is (or is close to) the corresponding weight average I_{sp}. The small discrepancy which exists probably is due to the drag law approximation leading to Eq. (10).

Table 8. Summary of effective particle size calculations

Particle size distribution		$\epsilon = A/A_t$	r_p, μ		
r_{pi}	x_{pi}		From Eq. (10)	From wt average I_{sp}	Wt average particle size
0.5	0.1				
1.0	0.2				
1.5	0.3	9.79	1.54	1.62	1.80
2.0	0.2				
3.0	0.1				
4.0	0.1				
0.5	0.1				
1.0	0.2				
1.5	0.3	20.09	1.49	1.53	1.80
2.0	0.2				
3.0	0.1				
4.0	0.1				
0.5	0.2				
1.0	0.3				
1.5	0.2	9.79	1.30	1.38	1.60
2.0	0.1				
3.0	0.1				
4.0	0.1				
0.5	0.4				
1.0	0.3				
1.5	0.1	9.79	0.98	1.01	1.20
2.0	0.1				
3.0	0.05				
4.0	0.05				
0.5	0.1				
1.0	0.1				
1.5	0.1	9.79	1.97	2.13	2.40
2.0	0.2				
3.0	0.3				
4.0	0.2				

References

1. Sehgal, R., and Strand, L., A *Theory of Low-Frequency Combustion Instability in Solid Rocket Motors,* Technical Memorandum No. 33-130, Jet Propulsion Laboratory, Pasadena, California, May 1, 1963.

2. Anderson, F. A., Strand, L. D., and Strehlow, R. A., *An Experimental Investigation of the Low-Pressure Combustion Limits of Some Solid Propellants,* Technical Memorandum No 33-134, Jet Propulsion Laboratory, Pasadena, California, June 3, 1963, (Confidential).

3. Anderson, F. A., Strehlow, R. A., and Strand, L. D., *Low Pressure Rocket Extinction,* Technical Report No. 32-509, Jet Propulsion Laboratory, Pasadena, California, January 17, 1964.

4. Price, E. W., *Low-Frequency Combustion Instability of Solid Rocket Propellants, 1 July–1 September 1962,* Technical Progress Report 301, NOTS TP 3107, U.S. Naval Ordnance Test Station, China Lake, California, December 1962.

5. von Elbe, G., "Theory of Solid Propellant Ignition and Response to Pressure Transients," *Bulletin of the Interagency Solid Propulsion Meeting July 1963,* CPIA Bulletin No. 18, Chemical Propulsion Information Agency, John Hopkins University Applied Physics Laboratory, Vol. III, pp. 95–127, June 1963. (Volume–Confidential, article–unclassified).

6. Robillard, G., and Lenoir, J. M., "The Development of a New Erosive Burning Law," Publication No. 91, Jet Propulsion Laboratory, Pasadena, California, April 1, 1957.

7. Montgomery, L. C., "Sterilization of Solid Propellants," SPS *37-20,* Vol. IV, pp. 58–60, Jet Propulsion Laboratory, Pasadena, California, April 30, 1963.

8. Kliegel, J. R., "One-Dimensional Flow of a Gas-Particle System," presented at the IAS 28th Annual Meeting, New York City, N.Y., January 25–27, 1960.

9. Ingham, J. D., et al., *Relative Performance of High Energy Propellants and the Importance of the Weight Fraction of Condensables in the Exhaust,* Technical Memorandum No. 33-6, Jet Propulsion Laboratory, Pasadena, California, March 21, 1960.

IX. Polymer Research

A. The F^{19} NMR of Oxypropyl Trifluoroacetates[1]

S. L. Manatt, J. D. Ingham, N. S. Rapp, and D. D. Lawson

The molecular structure of polyoxyalkylene glycols can profoundly affect the processing and stability of polyurethanes prepared from them. Of immediate concern are the effects on chemical properties resulting from different types of end groups, e.g., whether they are primary or secondary hydroxyl, and any effects of differences in stereoisomerism.

Recent work (Ref. 1) has shown that, because of the large number of electrons associated with the F^{19} nucleus, chemical shifts between trifluoroacetates of primary and secondary hydroxyls are of the order of 10–15 cps, with the primary trifluoroacetates at lower field. Corresponding shifts in the proton magnetic resonances between primary and secondary acetates are observed; however, these are only 1–6 cps. Therefore, F^{19} nuclear magnetic

---

[1] Joint contribution with the Chemistry Section, Space Sciences Division.

resonance (NMR) is being used to establish the structure of terminal hydroxyls in polyoxypropylene glycol (PPG).

1. F^{19} Results for Trifluoroacetates of PPG-425, PPG-2000, and PPG-900 Terminated with One Hydroxyl

The F^{19} NMR spectra of the trifluoroacetates of both PPG-425 and PPG-2000 show the presence of little or no primary hydroxyl. However, the resonance at high magnetic field, which is attributed to secondary hydroxyl, shows a doublet structure. The high field member of the doublet is of slightly greater intensity than that at low field and the difference in relative intensity of these doublets is greater for PPG-425 than for PPG-2000. Since no other perturbation appears likely, the doublet indicates *two different types of secondary hydroxyl in PPG.* We felt that a polyoxypropylene glycol with one end capped with an alkoxy group should provide some information on the origin of this nonequivalence. Therefore, propylene oxide was polymerized by initiation with sodium n-propoxide and gave a polyoxypropylene of number-average molecular weight 900 with a propoxyl group at one end and a secondary hydroxyl group at

Fig. 1. F¹⁹ NMR spectra of the trifluoroacetate of polyoxypropyl alcohol

the opposite end of each chain. The F^{19} NMR of the trifluoroacetate of this polymer was a doublet with a chemical shift for secondary trifluoroacetate and with each member of equal intensity, thus indicating approximately equal numbers of molecules with each kind of secondary hydroxyl (Fig. 1).

2. Discussion of F^{19} NMR Results

In order to account for two types of secondary hydroxyl in PPG, two tentative explanations have been proposed (Ref. 1). The first is a difference in positional isomerism of methyl group of the propoxy unit adjacent to the terminal unit, that is, Structures I and II:

$$\underset{\text{I}}{CH_3\overset{OH}{\underset{|}{C}}HCH_2OCH_2\overset{CH_3}{\underset{|}{C}}HO\sim} \qquad \underset{\text{II}}{CH_3\overset{OH}{\underset{|}{C}}HCH_2O\overset{CH_3}{\underset{|}{C}}HCH_2O\sim}$$
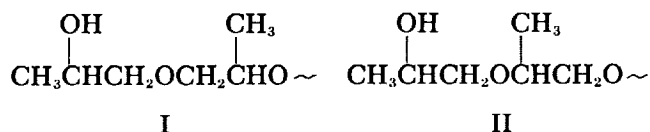
Alternatively, there could be a difference in stereoisomerism of the adjacent asymmetric carbons of the two propoxy units at the chain ends:
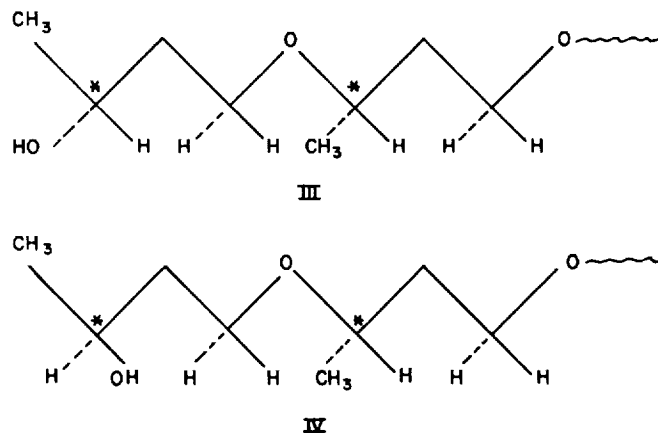
Structures I and II imply that the mechanism of polymerization is the origin of the observed nonequivalence of the secondary trifluoroacetate groups via the following reaction steps:

$$\underset{}{CH_3\overset{OH}{\underset{|}{C}}H\overset{O^-}{\underset{|}{C}}H_2} + CH_3\overset{O}{\overset{\diagup\diagdown}{CH}}CH_2 \rightarrow CH_3\overset{OH}{\underset{|}{C}}HCH_2OCH_2\overset{CH_3}{\underset{|}{C}}HO^-$$

$$V$$

$$V \xrightarrow[+ \text{ TERMINATION}]{CH_3\overset{O}{\overset{\diagup\diagdown}{CH}}CH_2}$$

$$\underset{\text{I}}{CH_3\overset{OH}{\underset{|}{C}}HCH_2OCH_2\overset{CH_3}{\underset{|}{C}}HO}\sim\sim\sim\sim \underset{\text{II}}{CH_2\overset{CH_3}{\underset{|}{C}}HOCH_2\overset{CH_3}{\underset{|}{C}}HOH}$$

$$VI$$
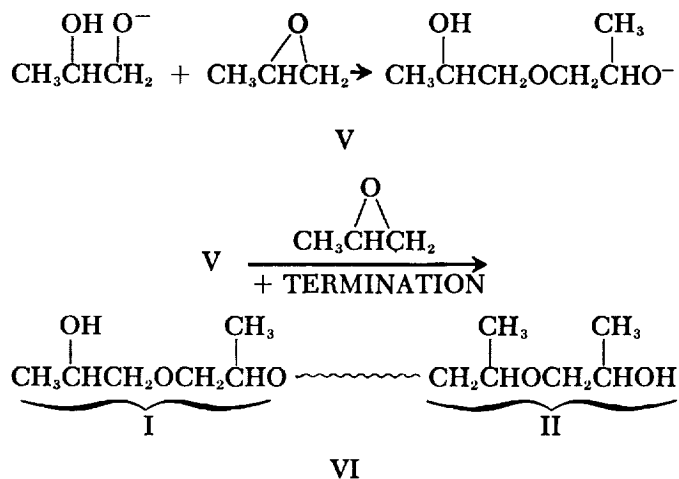
If the initiating anion used is from a monofunctional alcohol RO^- instead of $CH_3CHOHCH_2O^-$, only end groups of Type II should be formed. Because the monofunctional initiator gave a doublet, either the above mechanism is incorrect or the doublet arises because of a difference in stereoisomerism of the end groups. However, any alternative mechanism that would give both End Groups I or II on different molecules when monofunctionally initiated does not seem reasonable. From existing evidence then, it appears that the doublet is produced by differences in the chemical shifts of the trifluoroacetates of Structures III and IV. Further investi-

gation will include F¹⁹ NMR measurements on compounds of known stereochemical configuration to unequivocally prove or disprove this tentative conclusion.

B. Isolation, Identification, and Synthesis of Dipropylene Glycol Isomers II. 2,4-Dimethyl-3-Oxapentane-1,5 Diol

D. D. Lawson

In Ref. 2 the chemical characterization of 4-oxaheptane-2,6 diol (disecondary dipropylene glycol) was reported. As a continuing effort the synthesis, isolation, and physical properties of the 2,4-dimethyl-3-oxapentane-1,5-diol (diprimary dipropylene glycol) have been carried out. In particular, the preparation of the diastereomers and

configuration relative to dilactic acid have been determined along with an alternate synthesis of the diprimary dipropylene glycols. The gas-liquid partition chromatography (GLPC) data over a temperature range on a Carbowax 20M column and the separation of the diastereomers is also noted.

In connection with the study of lactic acid metabolism, Pierre Vieles and coworkers (Ref. 3) prepared a large number of derivatives of dilactic acid (I) and determined their configurations relative to lactic acid. In the case of diethyl dilactate (II) prepared from the sodium salt of ethyl lactate and ethyl α-bromopropionate, the ratio of dl pairs to meso form was five to one. The dilactate esters were separated by careful fractional distillation to obtain the lower boiling dl form. The meso ester remained in the residue and was further purified by preparative gas chromatography. To determine which ester represents the meso and which the dl form, the diamide (III) was prepared by heating the appropriate ester (obtained from preparative GLPC) with ammonia in a sealed tube. On reduction of either dl or meso diethyl dilactate (II) to the diprimary glycol (IV) with
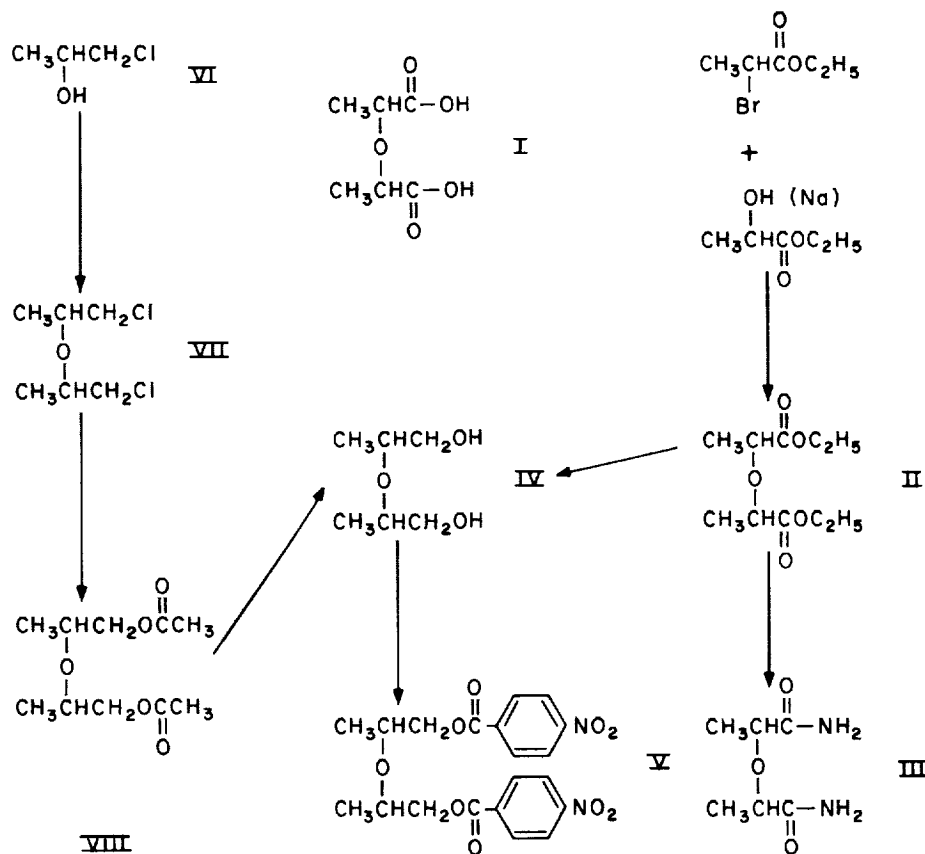
Fig. 2. Synthetic scheme for 2,4-dimethyl-3-oxapentane-1,5-diol

Table 1. Properties of diastereomers of 2,4-dimethyl-3-oxapentane-1,5-diol and related substances

Compound	Melting point, °C	Boiling point, °C/mm	Index of refraction (t, °C)	Density (t, °C)	Molecular formula	Carbon, % Calculated	Carbon, % Found	Hydrogen, % Calculated	Hydrogen, % Found	Nitrogen or chlorine, % Calculated	Nitrogen or chlorine, % Found
IV 2,4-dimethyl-3-oxapentane-1,5-diol											
(dl)		87/0.15	1.4426 (25) 1.4322 (60)		$C_6H_{14}O_3$	53.71	53.52	10.52	9.90		
(meso)			1.4356 (25)				53.46		10.61		
V 1,5-bis (p-nitrobenzoyloxy)-2,4-dimethyl-3-oxapentane											
(dl)	141-2				$C_{20}H_{20}N_2O_9$	55.55	55.57	4.66	4.59	6.48 (N)	6.34 (N)
(meso)	115-6						55.06		4.61		6.63
II 2,4-dicarbethoxy-3-oxapentane											
(dl)		118–120/20	1.4143 (28.1)	1.023 (28.1/25)	$C_{10}H_{18}O_5$	55.03	55.01	8.31	8.20		
(meso)			1.4193 (28.1)				54.79		8.35		
III 2,4-dicarbamyl-3-oxapentane											
(dl)	182-3				$C_6H_{12}O_3N_2$	44.99	45.08	7.55	7.95	17.49 (N)	17.53 (N)
(meso)	133-4						45.76		7.84		17.92
VII bis-(2-chloroisopropyl)-ether		181/731 89–90/28	1.4472 (20)	1.111 (20/25)	$C_6H_{12}OCl_2$	42.12	42.19	7.07	7.25	41.45 (Cl)	4096 (Cl)
VIII 1,5-diacetoxy-2,4-dimethyl-3-oxapentane		118/9	1.4243 (20)	1.041 (20/25)	$C_{10}H_{18}O_3$	55.03	54.98	8.31	8.40		

lithium aluminum hydride and dry diethyl ether no apparent epimerization occurs. The course of the reduction was verified by GLPC, using a Carbowax 20M column which can easily separate the esters or glycols. The bis-p-nitrobenzoate derivatives (V) of the glycols were prepared by heating p-nitrobenzoyl chloride with either the meso or dl form of IV and then recrystallizing the diesters from ethanol. It is apparent from the melting points of the bis-p-nitrobenzoates that Britton and Sexton (Ref. 4) prepared only the dl-2,4-dimethyl-3-oxapentane-1,5 diol (IV).

An alternate synthesis of IV is to heat in an autoclave the bis-(2-chloroisopropyl)-ether (VII) with potassium acetate and acetic acid as a solvent to give the diacetate (VIII) in good yield. The dichloro-ether was readily obtained by dehydration of 1-chloropropanol-2 (VI) with sulfuric acid. The 2-chloropropanol-1 will not yield an ether under similar conditions (Ref. 5). An ester interchange between methanol and VIII, or saponification with barium hydroxide, gave a mixture of the diastereomers of IV, which then could be separated by preparative GLPC. Table 1 lists the physical characteristics and analytical data for the above compounds. Fig. 2 shows

the synthetic scheme and structure of the materials. In Fig. 3 a plot is given of the specific retention volume (Vg^{760}) of the dl and meso forms of 2,4 dimethyl-3-oxapentane-1,5 diol versus $1/T$ with propylene glycol being included as a reference.

Further work on dipropylene glycol will be directed toward the isolation, identification, and synthesis of 2-methyl-3-oxahexane-1,5 diol (primary-secondary dipropylene glycol).

C. Structure of Poly-9-Vinyl-Anthracene

A. Rembaum and A. Henry

1. Introduction

In a previous report (Ref. 6) evidence was presented which indicated that 9-vinylanthracene (VA) does not
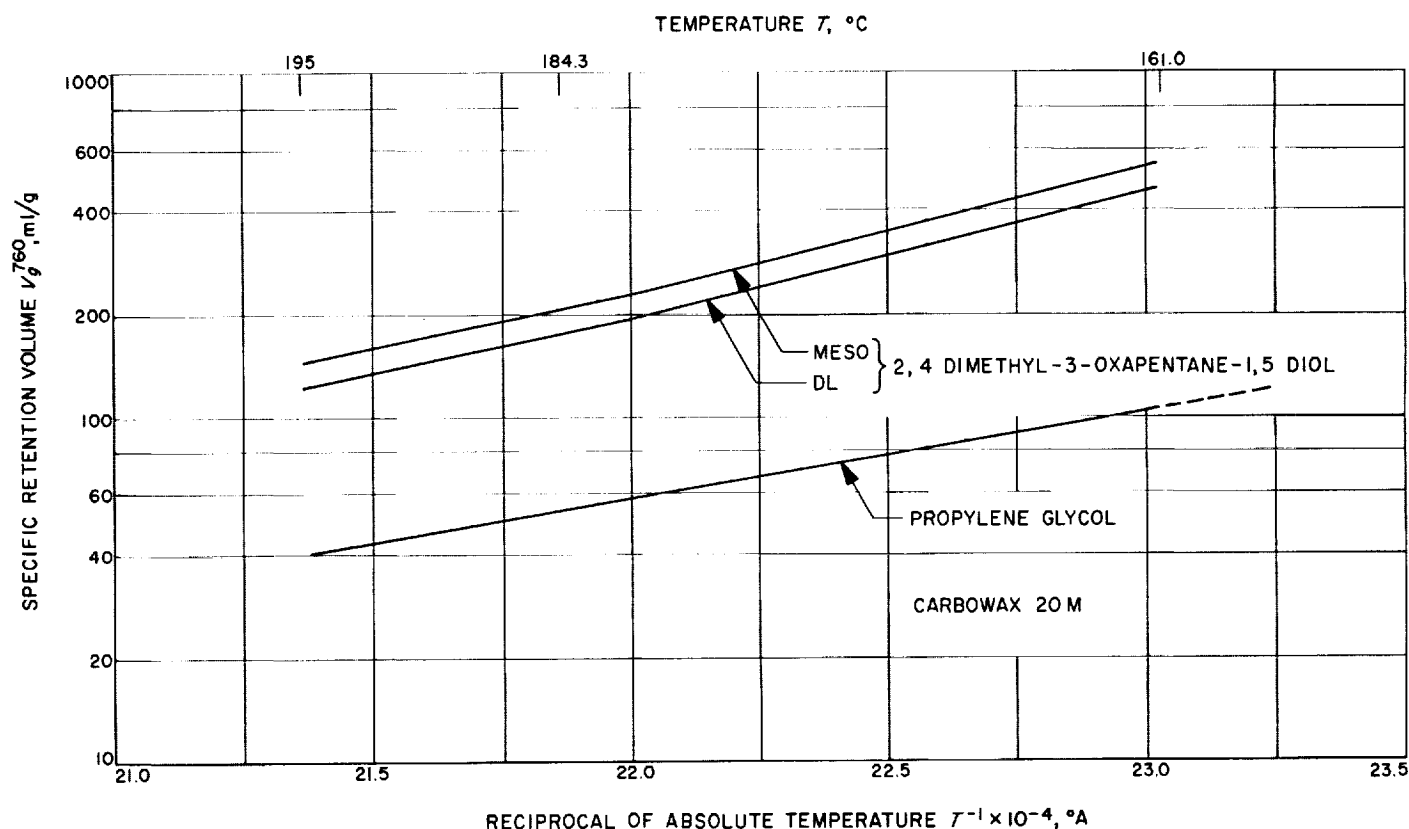
Fig. 3. Specific retention volume vs reciprocal temperature for glycols

polymerize in the conventional way by 1,2 addition and that chain extension proceeds across the middle anthracene ring, i.e., by 1,6 addition leading to 1,1-diphenylethylene (DPE) type of structures. Only low molecular weight polymers could be obtained and the number-average molecular weight of poly-9-vinylanthracene (PVA) was found to be of the order 1500–2000. These results were recently confirmed by independent research workers (Ref. 7).

From further investigations the following significant facts emerged:

(1) Fractionation of PVA by means of column chromatography permits the isolation of a polymer fraction, the molecular weight of which is of the order of 5000.

(2) Additional evidence based on spectroscopic results, as well as on the reaction between PVA and maleic anhydride, leaves little doubt that the previous postulate of 1,6-addition polymerization was correct. A comparison of results obtained by three different analytical methods yielded a quantitative estimate of the concentration of anthracene rings in the polymer.

2. Results

a. Polymerization. The experimental techniques for the polymerization under high vacuum and determination of polymer yields using sodium naphthalene initiator were previously described (Ref. 8). Additional data on this polymerization system where the initiator was replaced by butyllithium are shown in Table 2.

Table 3. Fractionation of PVA

Fraction	Weight %	M_n^c	M_n^b	Extinction coefficient at 258 $m\mu$ $lg^{-1} cm^{-1}$
1	35.50	1050	850	202
2	18.55	1700	1450	170
3	38.20	3300	2800	189
4	3.98	4200	—	124
5	3.77	5300	—	141
	100.00			

*Number average molecular weight determined by means of a Machrolab osmometer.
*Number average molecular weight determined by ebulliometry.

Table 4. Fractionation of PVA*

Fraction	Weight %	M_n	Extinction coefficient at 258 $m\mu$ $lg^{-1} cm^{-1}$
1	27.0	800	210
2	34.1	2400	150
3	9.9	2540	120
4	28.9	4500	113
5	0.1	5800	140
	100.0		

*Original M_n (number average molecular weight: 1800.

Table 2. Polymerization of VA by means of butyllithium

Butyllithium, moles/l	VA, moles/l	Temperature, °C	Time, hr	Yield, weight %	Molecular weight*	Extinction coefficient at 258 $m\mu$, $lg^{-1} cm^{-1}$
8×10^{-3}	0.4	−80	2	17.7	1750	122
8×10^{-3}	0.4	0	2	28.0	—	140
8×10^{-3}	0.4	50	2	79.4	2370	180
8×10^{-3}	0.4	100	2	73.7	1650	185
5×10^{-3}	2.5	100	5	99.8	1100	198
1.6×10^{-3}	0.8	100	5	85.8	1300	170
0.6×10^{-3}	0.3	100	5	56.2	1350	167
1.6×10^{-3}	0.4	50	2	85.0	1800	181
2×10^{-3}	0.4	50	2	87.7	1350	185
4×10^{-3}	0.4	50	2	94.5	1080	187
1×10^{-1}	0.4	50	2	96.0	900	200

*Number-average molecular weight determined by means of a Mechrolab osmometer.

b. Fractionation. Two polymer samples prepared by means of butyllithium were fractionated separately on an alumina column, using n-hexane-benzene mixtures as eluents (10-90 and 50-50 volume %) at room temperature. The highest molecular weight fractions were obtained by elution with pure benzene. The results are shown in Tables 3 and 4.

c. Spectroscopic studies. VA as well as PVA yields colored solutions when reacted with sodium in the absence of air. The visible spectrum of the monomer (Curve 2, Fig. 4) differs from the visible spectrum of the polymer (Curve 1, Fig. 4). The latter exhibits an extra absorption peak at 460–480 mμ, i.e., at the same wavelength at which 1,1-diphenylethylene (DPE) absorbs under identical conditions.

In Fig. 5 are shown spectra in the visible wavelength range of synthetic mixtures of 9-ethyl anthracene (EA) and DPE reacted with sodium. These are characterized by the same peak at 460–480 mμ. The changes of optical density of DPE with time have been observed previously (Ref. 9) and are attributed to the establishment of equilibria between DPE and DPE radical ions.

The ultraviolet spectra of DPE, PVA, and synthetic mixtures of DPE and VA are recorded in Fig. 6.

Fig. 4. Visible spectra of → poly-9-vinylanthracene (PVA) on sodium and → 9-vinylanthracene (VA) on sodium tetrahydrofuran
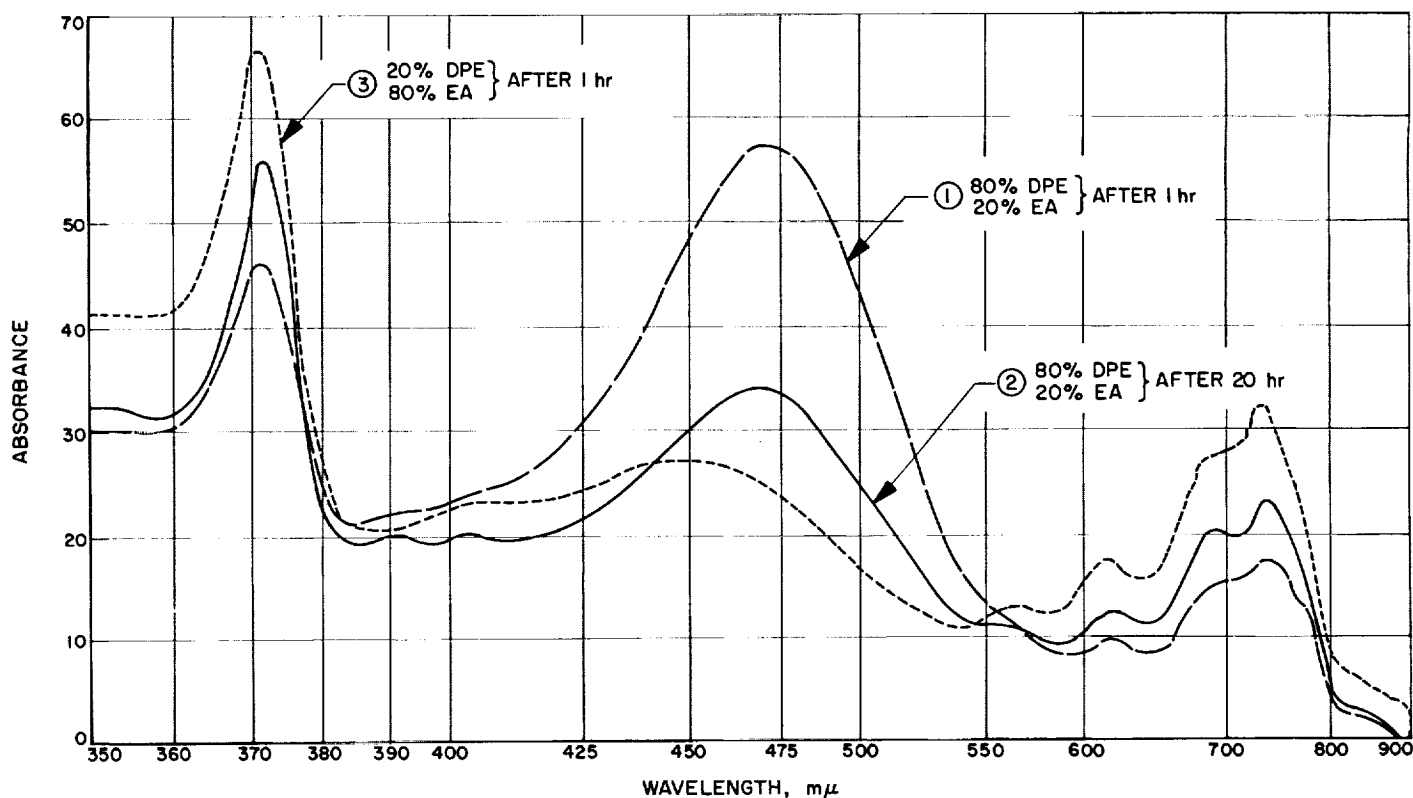
Fig. 5. Visible spectra of synthetic mixtures of 1,1-diphenylethylene (DPE) and 9-ethyl anthracene (EA) reacted with sodium in tetrahydrofuran
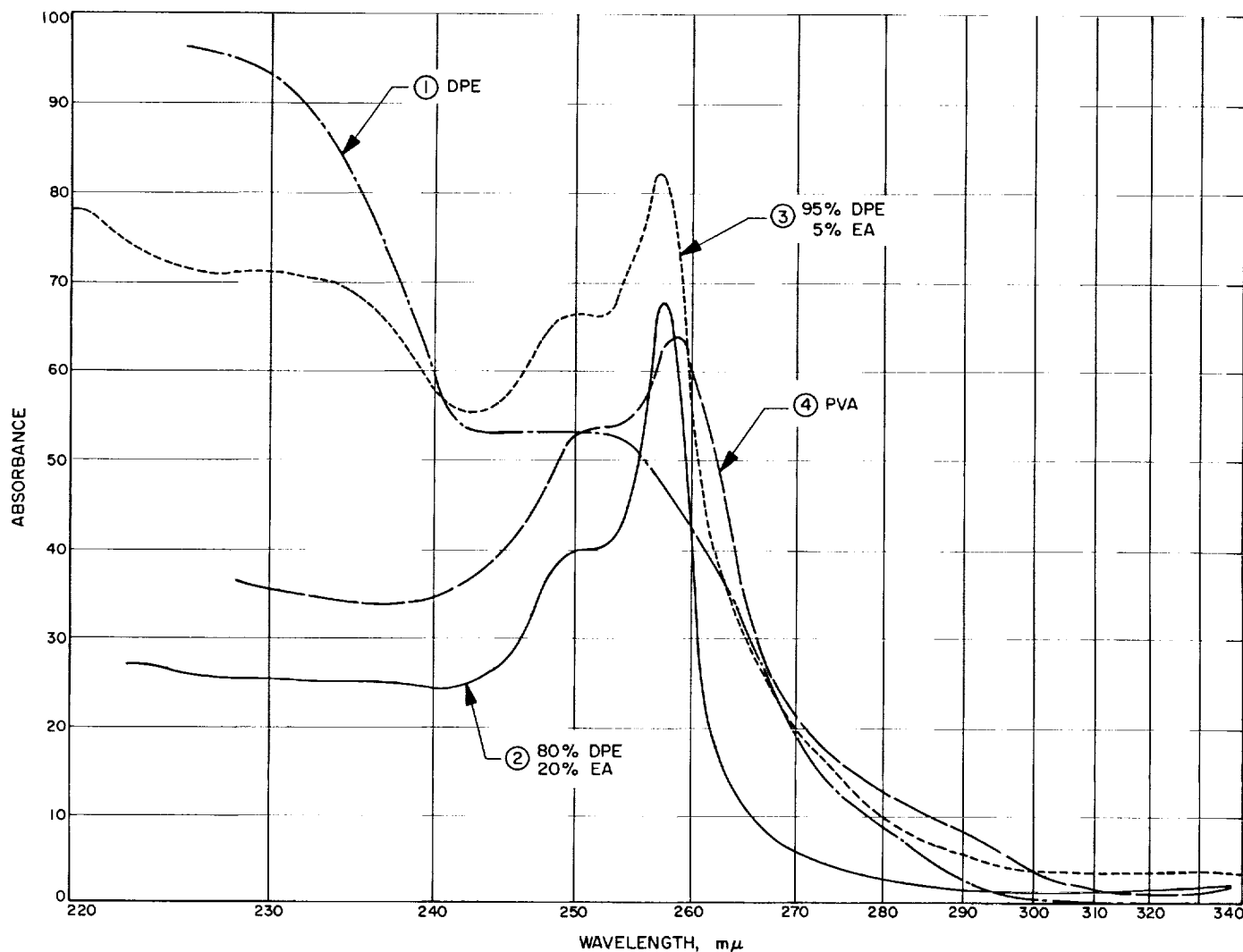
Fig. 6. Ultraviolet spectra of 1,1-diphenylethylene (DPE), 9-ethyl anthracene (EA), and poly-9-vinylanthracene (PVA) in dimethoxyethane

The examination of the infrared spectra of EA and PVA yielded additional evidence for a 1,6 type of polymerization. In Fig. 7 the absorption spectrum of PVA and EA (identical concentration) are compared in the range of 9 to 20 μ. The reduction in intensity of the peak at 13.7 μ is obviously due to a decrease of anthracene structures in the polymer. The ratio of extinction coefficients of the monomeric moiety to the polymer at 13.7 μ gives the amount of anthracene structures present after polymerization (Table 5).

d. Reaction with maleic anhydride. The formation of the maleic anhydride adduct was used as an independent check of the amount of anthracene moiety present in PVA. The reaction was carried out under conditions described

in the literature (Ref. 10), and the carbon and hydrogen content of the isolated adducts was determined. The weight percent of anthracene structures calculated from the oxygen content in the reaction product is recorded in Table 6.

3. Conclusions

The fractionation of PVA and isolation of polymer fractions of a degree of polymerization of 25 and the spectroscopic results offer a conclusive proof for the successive addition of monomer units across anthracene rings. Furthermore, construction of Fisher-Hirshfelder models (Figs. 8 and 9) show that 1,6-type polymerization leads to structures considerably less hindered than those
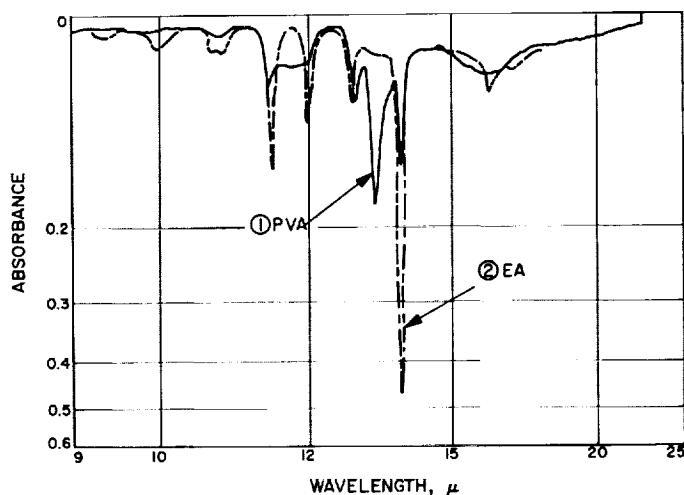
Fig. 7. Infrared spectra of poly-9-vinylanthracene (PVA) and 9-ethyl anthracene (EA) in CS_2

Fig. 8. Molecular model representing 1,6 type of polymerization of 9-vinylanthracene

formed by 1,2 addition. The molecular model shown in Fig. 9 represents a dimer formed by reaction of VA through the vinyl group. Addition of a third monomer unit by the 1,2 mechanism appears to be physically impossible due to the steric interference of hydrogen atoms, while formation of a long polymer chain by reaction across the middle ring of the anthracene molecule (Fig. 8) may take place with little steric hindrance.

The DPE postulate seems at first to be inconsistent with the ultraviolet spectrum of the polymer, since the

Table 5. Infrared analysis of PVA

Sample	Number-average molecular weight	Weight % anthracene structures
PVA	1100	31.0
PVA	2100	24.4
PVA	2860	32.7
PVA	3600	32.3

Fig. 9. Molecular model representing 1,2 type of polymerization of 9-vinylanthracene

Table 6. Analysis of maleic anhydride-PVA adducts

% C	% H	% O[a]	Weight % anthracene structures
89.41	5.61	4.98	30.9
89.95	5.60	4.45	27.6
88.52	5.50	5.98	37.1
88.82	5.70	5.48	34.0
[a]Oxygen by difference.			

latter does not exhibit the plateau shown by DPE in the wavelength range of 242 to 254 $m\mu$ (compare Curve 1 with Curve 4 in Fig. 6). This is not surprising since the experimental oscillator strength of EA is very much greater than that of DPE; and this is confirmed by a spectroscopic examination of synthetic mixtures. It is clear from Fig. 6 that EA remains the predominant species in the spectrum up to 95% of DPE, and even at this high concentration the plateau of DPE is not apparent.

The quantitative analysis by means of three independent methods (infrared, ultraviolet, and reaction with maleic anhydride) leads to the conclusion that the maximum concentration of anthracene rings in the polymer is about 35%. The ultraviolet method yielded values approximately 10% lower (Ref. 6) than those obtained from the infrared or maleic anhydride analysis. This discrepancy may be due to the fact that maleic anhydride is known to react with DPE (Ref. 9) and the structure resulting from 1,6 type of addition polymerization of VA is likely to contribute to the absorption intensity at 13.7 μ. In both cases, therefore, the results would be too high and for this reason the ultraviolet analysis is considered to be the most accurate.

References

1. Rapp, N. S., Hardy, J., Manatt, S. L., and Ingham, J. D., "Identification of Primary and Secondary Hydroxyl Groups in Propylene Glycol and Polyoxypropylene Glycols by F^{19} NMR of their Trifluoroacetates," SPS 37-24, Vol. IV, pp. 94–98, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.

2. Lawson, D. D., and Havlik, A. J., "Isolation, Identification, and Synthesis of Dipropylene Glycol Isomers I. 1, 1'-Oxydipropanol-2," SPS 37-22, Vol. IV, pp. 94–96, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.

3. Vieler, P., "L'Acide Diladylique Brut," Annals de Chemie II, Vol. 3, p. 143, 1935.

4. Sexton, A. R., and Britton, E. C., "Synthesis and Identification of Dipropylene Glycol," Journal of the American Chemical Society, Vol. 75, p. 4357, 1953.

5. Dewael, A.,"Sur Quelques Ether-Oxyde Deverant de la Chlorhydine Propyleneque," Bulletin of the Belgian Chemical Society, Vol. 34, p. 343, 1925.

6. Rembaum, A., Henry, A., and Waits, H., "Anionic Polymerization of 9-Vinyl Anthracene and Semiconductivity of Poly-9-Vinylanthracene Iodine Complexes," SPS 37-19, Vol. IV, pp. 103–110, Jet Propulsion Laboratory, Pasadena, California.

7. Michel, R. H., and Baker, P., "The Structure of Anionically Prepared Poly-9-vinylanthracene," Journal of Polymer Science, Part B, Vol. 2, pp. 163–169, 1964.

8. Tobolsky, A. V., Rembaum, A., and Eisenberg, A., "Equilibrium Polymerization of α-methyl Styrene," Journal of Polymer Science, Vol. 45, p. 347, 1960.

9. Jagur, J., Monteiro, H., and Szwarc, M., "Chemistry of Radical Ions," Transactions of the Faraday Society, No. 486, Vol. 59, p. 1353, 1963.

10. Kloetzel, M. C., "The Diels Alder Reaction with Maleic Anhydride," pp. 28–35, Organic Reactions, John Wiley & Sons, New York City, N.Y., 1948.

X. Propulsion Research

A. Gas Side Boundary Phenomena

R. W. Rowley

1. Introduction

An understanding of the thermal, chemical, and fluid-mechanical processes which constitute combustion in a liquid rocket engine has long been the goal of the designer. Such understanding has proven so difficult to acquire, however, that, in general, it is necessary to accept grossly simplified models of complex processes for design purposes and then use "cut and try" techniques during the development of an engine. A case in point is the transfer of heat from the hot combustion gases to the thrust chamber wall. At present it is necessary to assume an all-gas, well-mixed environment within the combustion chamber from which heat is transferred to the boundaries primarily by convection. This allows the use of available semiempirical or boundary layer analyses such as those of Bartz (Refs. 1 and 2). However, the most widely used spray combustion techniques, wherein combustion takes place in a moving field of droplets, do not completely satisfy these assumptions. The flow field is obviously not all gaseous, and droplets may persist as far downstream as the nozzle throat. Impingement of droplets on the thrust chamber wall and the resultant liquid film, however transitory, may be expected to modify the pure forced convection heat transfer process. In addition, large gradients in local mass flow rate and local mixture ratio exist in droplet fields produced by the most popular injection schemes. These gradients in turn result in corresponding gradients in the gaseous combustion products. Thus, any understanding of the processes taking place at the wall of a real rocket engine must consider local conditions rather than the mean-mixed ideal.

Previous studies at this Laboratory (Refs. 3 and 4) have demonstrated the existence of nonuniform heat transfer rates to thrust chamber walls. These investigations were conducted with unlike impinging stream injectors having controlled stream characteristics. Although reproducible local heat flux distributions were obtained and these distributions were obviously related to injector characteristics (Ref. 3), a useable correlation has yet to be devised. In Ref. 4 the nonuniform circumferential distribution of heat flux produced by a particular injector was shown to be grossly related to erosion in the throat

of an ablative thrust chamber. Data were also presented showing that the mixture ratio (i.e., chemical composition) of the flow near the wall was related to local chemical erosion of nozzle throat materials.

Results of these studies, demonstrating the existence of nonuniform conditions at the wall of an operating liquid rocket engine as a result of the nonuniform character of the combustion-flow processes associated with practical injection methods and thrust chamber geometries, suggested that a detailed investigation of injector-related wall phenomena could be both fruitful and useful. Accordingly, a long-term research investigation entitled "Gas Side Boundary Phenomena" has been initiated.

2. Experimental Apparatus

Initial experiments being conducted as part of this project are aimed at elucidating results of earlier programs and defining the physical processes which control heat transfer to surfaces exposed to a spray combustion flow field. Starting with a typical liquid rocket engine, a series of simplifications have been made in arriving at an apparatus which, although it no longer resembles a rocket engine, incorporates the salient characteristics of rocket engine combustion. These simplifications were included as a result of experience gained in previous programs where the complex geometry associated with the spray striking the thrust chamber wall and the complete lack of visual observation have hampered a real understanding of the phenomena which control heat transfer and erosion.

The device which was constructed for this initial phase consisted simply of an instrumented flat plate which could be inserted into a burning spray. No combustion chamber was used and no effort was made to exclude the surrounding air from the combustion process. Although some disadvantages might result from the aspiration of air into the droplet field, it was felt that the advantages of a free choice of orientation of the flat plate heat transfer surface and the ease of observation of the burning spray and its impingement on the flat plate more than compensated for the unknown effect of the air on the combustion process. As an additional simplification, the multi-element construction of typical injectors was abandoned in favor of a single injector element, thus avoiding complex interactions between the sprays formed by adjacent elements. The element chosen consisted of a jet of fuel impinging on a jet of oxidizer, i.e., an "unlike-impinging-doublet" element. The general configuration of the spray thus formed is shown in Fig. 1.

The fuel and oxidizer jets were formed by long, smooth bore tubes with sharp, burr-free exits. The tubes were 100-bore diameters long and the flow velocity was high enough to insure a fully developed turbulent velocity profile. Reynolds number was about 100,000 for both jets. The jets were formed in this manner in order to insure stable, reproducible hydraulic characteristics, a prerequisite to a reproducible spray. The characteristics of jets of this type are discussed in Ref. 5.

Arguments have been presented in Refs. 6 and 7 that impingement of jets of highly reactive liquids such as nitrogen tetroxide and hydrazine does not produce a field of bipropellant droplets but, rather, the streams "blow apart" as a result of gas evolution at the fuel-oxidizer interface. Since this phenomenon has not yet been adequately explored, and since some information on the properties of sprays produced with nonreacting liquids is available (Ref. 8), the present investigation is being conducted with propellants which exhibit negligible liquid-phase reaction. The oxidizer being used is inhibited white fuming nitric acid (WIFNA) and the fuel is methanol. When combined, these propellants do exhibit some temperature increase; however, it is not felt that the evaporation rate is increased sufficiently to appreciably disturb the mixing process. The spray produced with WIFNA/methanol would thus be expected to exhibit mass and mixture ratio distributions similar to those obtained with nonreacting fluids having similar physical properties. Such information is an obvious prerequisite to an understanding of local conditions when the spray impinges on the wall and is much more easily obtained by using nonreacting liquids than by using the actual propellants. These propellants also have the advantage of producing relatively simple equilibrium reaction products.

Ignition of the spray was accomplished with a coil of Nichrome wire which was heated electrically to a cherry red. Once the spray was ignited the reaction was self-sustaining, eliminating the need for a continuous external ignition source. The flame was similar in appearance to the flame produced by methanol burning in air except for an orange tinge, apparently the result of nitrogen oxides introduced by the oxidizer.

Since the combustion process was not confined, the gaseous products tended to rise as a result of the density difference between the hot gases and the surrounding ambient air. As the gases rose above the droplet field the energy they contained was not available for vaporizing the propellant droplets, and the rate of propellant consumption was thus reduced. At the flow rate which

PROPELLANT	OXIDIZER WIFNA	FUEL METHANOL
ORIFICE DIAMETER, in.	0.232	0.173
FLOW RATE, lb/sec	1.00	0.42
AVERAGE JET VELOCITY, ft/sec	36	52

RESULTANT MOMENTUM LINE

FUEL ORIFICE

30°

8°

30°

OXIDIZER ORIFICE

TOP VIEW

FLAT PLATE, 6 X 12 X 0.090 in. STAINLESS STEEL

PLATE MOVABLE

A

IMPINGEMENT POINT

FUEL   OXIDIZER

A

APPROXIMATE BOUNDARY OF SPRAY (90% OR MORE OF THE INJECTED MASS IS INCLUDED WITHIN THIS BOUNDARY)

SIDE VIEW

SECTION A

A

Fig. 1. Configuration of test apparatus

was used in the tests conducted to date combustion ceased about 6 ft from the impingement point, apparently as a consequence of the reduced heat transfer between products and droplets. A large number of unconsumed droplets were visible downstream of this point. This lack of confinement represents, perhaps, the most significant drawback of the present apparatus in that the flow is markedly different from that in a rocket engine where the gases are constrained to pass through the droplet field at high velocity before exiting through the nozzle, transferring heat to the droplets in the process.

3. Results

Some preliminary experiments have been conducted in which a stainless steel plate, 0.090 in. thick, to which thermocouples had been welded, was exposed to the burning spray. The electrical output of the thermocouples, which were attached to the rear surface of the

plate, was recorded during a series of tests in which the front surface of the plate was heated by the spray. The rate of temperature rise above the initial, ambient, temperature of about 70°F may be regarded as an indication of the heat transfer rate to the front surface of the plate. The surface of the plate was oriented perpendicular to the resultant momentum line of the spray with the resultant passing through the center of the plate as shown in Fig. 1. The temperature rise recorded by a thermocouple at this location, which is a stagnation region, is shown in Fig. 2 as a function of distance from the impingement point and time from initiation (and simultaneous ignition) of the spray. Thermocouples at other locations on the plate recorded essentially the same temperatures.

Although thermal equilibrium was not reached during these 10-sec tests, it would appear that a maximum temperature rise of 120–150°F might be expected at the
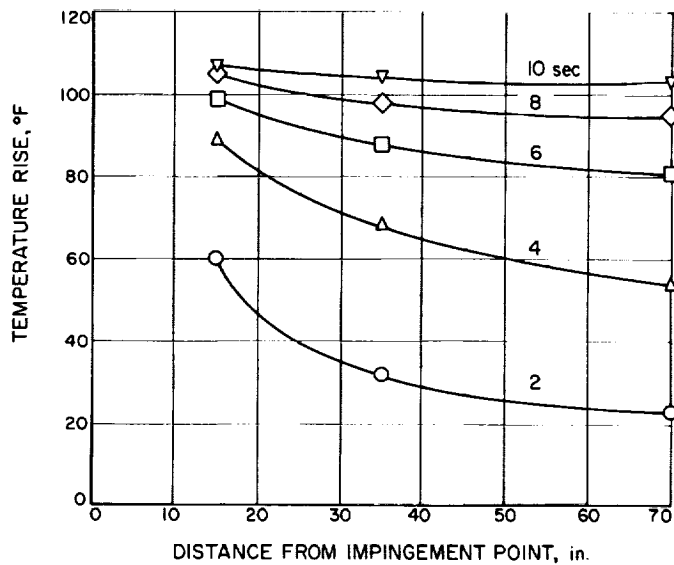
Fig. 2. Temperature rise at rear surface of flat plate

rear surface, corresponding to an equilibrium temperature of approximately 200°F. Since the plate was not cooled, the equilibrium temperatures of the front and rear surfaces would be expected to be approximately the same. This temperature is surprisingly low and virtually independent of distance from the impingement point. The data also show that early in the tests, i.e., when the temperature of the plate was low, the heat transfer rate (as indicated by the rate of temperature rise) increased as the plate was moved closer to the impingement point. The maximum heat transfer rate observed was of the order of 0.1 Btu/in²sec.

This indication that the heat transfer rate is a function of distance from the origin of the spray suggests a possible relationship to local mass flow rate of fluid approaching the plate. The divergence of the spray (shown schematically in Fig. 1) results in a reduction in local mass flow rate as the fluid travels away from the impingement point. The corresponding reduction in heat transfer to a flat plate is in agreement with convective heat transfer analyses such as those of Ref. 1, which assumes an almost first power dependence of heat transfer on mass flux in pipe flow, and with the data presented in Ref. 9, which more closely resembles the case under discussion. However, the energy level of the spray is quite low, as demonstrated by the low equilibrium temperature of the plate, and the heat transfer process is probably dominated by the large mass of unconsumed droplets rather than by convection from the combustion products.

These preliminary experiments have demonstrated that a simple spray can be produced which is capable of self-sustained reaction and which can provide information on spray combustion heat transfer. However, the present apparatus suffers from a serious deficiency in that the spray is not confined and the gaseous products rise out of the spray without contributing sufficient energy to the evaporation/combustion process. The droplets are thus consumed slowly and the flow approaching the plate is dominated by the liquid phase. The apparatus will eventually be modified to more closely approach the physical conditions of a rocket engine, where the droplets are (hopefully) completely consumed as they pass through the combustion chamber.

Additional experiments are planned with the device in its present form to define the local temperatures and heat transfer rates at the surface of the flat plate by using a more sophisticated technique in which thermocouples will be located at the front surface. High-speed movies of the liquid film and droplet impingement are also being made.

4. Propellant Chemistry

During the early development of the apparatus described, an anomalous chemical reaction was observed which, although not relevant to the data presented, is of some general interest. The injection device being used at that time consisted of two equal diameter, coaxially opposed orifices. Impingement of the fuel and oxidizer jets thus produced an axially symmetric flat sheet and, as a result, an axially symmetric spray fan. This symmetry greatly simplifies the intersection of the spray and a surface since the mass and mixture ratio distributions are now essentially two dimensional rather than three dimensional as is the case with a conventional doublet. However, there are two distinct disadvantages to this orifice geometry: (1) the injected mass is equally distributed around the circumference of the spray—thus local mass flow rates are lower than in the spray produced by a conventional doublet at the same total flow rate; and (2) carefully balanced jet hydraulic characteristics are required in order to insure that the resultant momentum is planar and that the impingement point is located midway between the ends of the orifices, rather than directly adjacent to the exit of one or the other orifice. With nitric acid/methanol propellants these hydraulic requirements specify a fuel-rich mixture ratio of 1.38.

Ignition was accomplished by inserting a small volume of hydrazine in the fuel line prior to the test such that

when the main propellant valves were opened, the hydrazine slug was forced out of the orifice by the methanol. The hypergolic reaction between hydrazine and nitric acid thus formed an ignition source for the methanol/nitric acid spray.

Following ignition by this technique, two different steady-state reactions were observed during a series of tests. One reaction produced the flame described previously, i.e., a faint yellow-orange flame having a billowy form as a result of the buoyancy of the hot gases. The second reaction produced a white cloud with no visible flame and no apparent buoyancy. The particles comprising the cloud appeared to move radially from the impingement point, along paths similar to those of the droplets in a spray produced with nonreacting liquids. This reaction, which appeared to be initiated at ignition, was produced continuously in tests as long as 8 sec, indicating that it was a self-sustaining phenomenon. A white solid, which collected during these tests on the apparatus that supported the orifices and is assumed to be the same material that formed the white cloud, was analyzed to be paraformaldehyde, which is a polymerized form of formaldehyde. There were no apparent differences in test conditions between tests which produced a flame and those which resulted in the formation of formaldehyde; the occurrence of one or the other result appeared to be random.

Thermochemical calculations show that the equilibrium products of the nitric acid/methanol reaction are primarily CO_2, N_2, and H_2O at stoichiometric conditions with added CO and H_2 at fuel-rich mixture ratios. However, formaldehyde is a probable intermediate product in the oxidation of methanol, suggesting that the appearance of paraformaldehyde was associated with "freezing" of the reaction at this partially reacted stage, while the appearance of a flame was an indication that the reaction had proceeded further toward completion. Some small change in operating conditions apparently determined whether or not the reaction could successfully drive itself through the formaldehyde intermediate.

A series of changes were made to the apparatus during these early stages of the program, both in an attempt to reliably produce the desired "flame" reaction and for other reasons. However, most of these modifications did not result in the consistent appearance of one reaction or the other. For example, the volume of the hydrazine starting slug was both increased and decreased by a factor of 2 from the volume initially chosen in order to correspondingly change the duration of the hypergolic

reaction used to ignite the methanol/acid spray. Changes were also made in the timing of the propellant valves in order to vary the relative time at which the fuel and oxidizer arrived at the impingement point. Both of these changes would be expected to affect the total energy released during the ignition phase as well as affecting the establishment of a recirculation pattern which probably feeds energy back to the liquid near the impingement point. However, neither change produced consistent results.

Subsequently, the oxidizer was changed from white fuming nitric acid to inhibited white fuming nitric acid, the latter containing 0.6% HF as a corrosion inhibitor, and the mixture ratio of the injected propellants was changed from fuel rich ($r = 1.38$), as required by hydraulic considerations, to stoichiometric ($r = 2.36$). Again, both reactions occurred in a random manner. This is particularly interesting since the change in mixture ratio, in addition to altering the concentration of reactants, also produced an imbalance in the relative momentum of the two jets which resulted in the impingement point being moved from a hydraulically balanced location midway between the ends of the two orifices to a point adjacent to the exit of the fuel orifice. However, any changes in the mixing process thus produced did not affect the nonreproducibility of the reaction.

Eventually the method of igniting the spray was changed from the hydrazine slug technique to a spark plug and a flame was produced consistently. The desired result having been achieved, the investigation of the mechanisms which caused the formaldehyde reaction was not pursued. However, since no other change was made to the system at the time the method of igniting the spray was changed, it would appear that the hydrazine slug technique in itself produced thermal or chemical conditions favorable for quenching the reaction.

At a later date the orifices were changed to a 60-deg included angle, rather than the opposed orientation, in order to increase the local mass flow rate in the region where the flat plate temperature measurements were made. This was considered necessary since, with the opposed jets, the flow field was completely dominated by the vertically moving (i.e., rising) buoyant combustion products rather than by drops or hot gases moving radially from the impingement zone. Although the 60-deg doublet produced a flow field more suitable for the flat plate experiments than that produced by the opposed jet orientation, the buoyancy of the combustion products remained a problem and, as mentioned previously, an

enclosed reaction zone appears desirable. Since the spark plug was readily shorted by the nitric acid, the ignition system was also changed in the interest of reliability, this time to the glow coil.

5. Summary

The apparatus currently being used, as described in Section 2, incorporates most of the simplicities originally desired, i.e., nonhypergolic propellants, geometrically simple heat transfer surfaces, ease of observation, and a relatively simple spray configuration. The change from the opposed jet orientation to the 60-deg doublet, with a resultant increase in complexity of the mass and mixture ratio distributions within the spray, represents the most significant degradation of the original concept. Although the device as it stands provides useful data on heat transfer rates to surfaces in spray combustion systems, the combustion process can be made to more closely approximate that in a rocket engine by confining the spray and the combustion products within a combustion chamber. Therefore, following a series of experiments with the present device using thermocouples mounted on the front surface of a plate, an enclosed apparatus will be constructed in which the flat plate experiments will be continued.

References

1. Bartz, D. R., "A Simple Equation for Rapid Estimation of Rocket Nozzle Convective Heat Transfer Coefficients," pp. 49–51, Jet Propulsion, January 1957.

2. Bartz, D. R., "An Approximate Solution of Compressible Turbulent Boundary-Layer Development and Convective Heat Transfer in Convergent-Divergent Nozzles," Transactions of the American Society of Mechanical Engineers, Vol. 77, No. 8, pp. 1235–1245, 1955.

3. Jaivin, G. I., Rupe, J. H., and Clayton, R. M., An Experimental Correlation of the Nonreactive Properties of Injection Schemes and Combustion Effects in a Liquid Propellant Rocket Engine: Part IV. Relating Heat Transfer to the Chamber Wall and the Injection Pattern, Technical Report, Jet Propulsion Laboratory, Pasadena, California, to be published.

4. Rowley, R. W., An Experimental Investigation of Uncooled Thrust Chamber Materials for Use in Storable Liquid Propellant Rocket Engines, Technical Report No. 32-561, Jet Propulsion Laboratory, Pasadena, California, February 15, 1964.

5. Rupe, J. H., On the Dynamic Characteristics of Free-Liquid Jets and a Partial Correlation with Orifice Geometry, Technical Report No. 32-207, Jet Propulsion Laboratory, Pasadena, California, January 15, 1962.

6. Elverum, Gerard W., Jr., and Staudhammer, Peter, The Effect of Rapid Liquid-Phase Reactions on Injector Design and Combustion in Rocket Motors, Progress Report No. 30-4, Jet Propulsion Laboratory, Pasadena, California, August 25, 1959.

7. Johnson, B. H., "Advanced Liquid Propulsion Systems—Injector Development," SPS 37-21, Vol. IV, pp. 109–112, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, June 30, 1963.

8. Rupe, J. H., The Liquid Phase Mixing of a Pair of Impinging Streams, Progress Report No. 20-195, Jet Propulsion Laboratory, Pasadena, California, August 6, 1953.

9. Gordon, Robert, and Cobonpue, John, "Heat Transfer Between a Flat Plate and Jets of Air Impinging on It," International Developments in Heat Transfer, American Society of Mechanical Engineers, Part II, pp. 454–460, 1961 (presented at the International Heat Transfer Conference).

XI. Advanced Propulsion Engineering

A. Liquid MHD Power Conversion

D. Elliott, D. Cerini, and D. O'Connor[1]

The long lifetimes required of electric-propulsion powerplants make nonrotating cycles attractive. A non-rotating powerplant under investigation at JPL is the liquid m··· tohydrodynamic (MHD) system shown schematically in Fig. 1. In this cycle a fluid, such as cesium, circulates in the vapor loop and causes a liquid metal, such as lithium, to circulate through an MHD generator in the liquid loop. The cesium leaves the radiator as condensate, flows through an electromagnetic pump to the mixer, vaporizes on contact with the lithium, atomizes and accelerates the lithium in the nozzle, separates from the lithium in the separator, and returns to the radiator. The lithium leaves the separator at high velocity (typically 500 ft/sec), decelerates through the production of electric power in the MHD generator, and leaves the
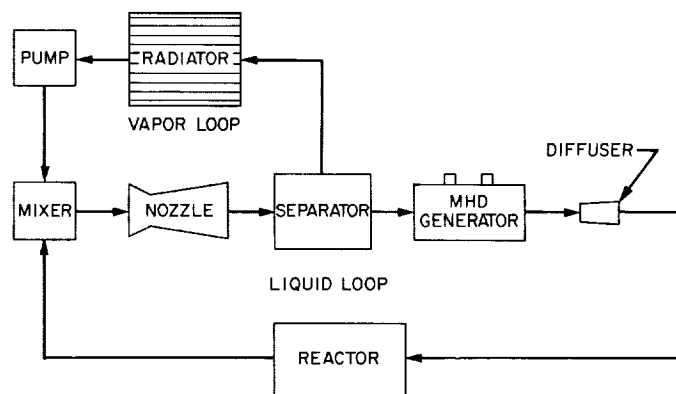
Fig. 1. Liquid MHD power conversion cycle

generator with sufficient velocity (typically 300 ft/sec) to return through a diffuser to the reactor where the lithium is reheated.

Results of nozzle, separator, and diffuser tests with nitrogen and water were reported in Refs. 1, 2, 3, and 4. Tests are now being conducted with liquid-metal MHD generators using NaK (78% potassium and 22% sodium) at ambient temperature as the working fluid. Subsequent

generator tests will employ two-phase flow to obtain higher efficiencies as discussed in Ref. 2. The purpose of the single-phase NaK tests is to compare theoretical predictions of generator performance with actual results for liquid-metal generators. The NaK generators are built to conform as closely as possible to the idealized configuration shown in Fig. 2, for which the theoretical performance can be readily computed.

1. Nomenclature and Assumptions

Referring to Fig. 2, the idealized MHD generator consists of a diverging rectangular duct of inlet width a_1, outlet width a_2, height b, and length L, within which the NaK contacts only the electrode faces. Insulated ducts extend a large distance upstream and downstream of the generator. A uniform magnetic field B_0 is applied over the length of the generator in the b direction. Upstream and downstream of the generator the field is zero. The liquid inlet velocity is V_1 and the electrode voltage produced is E. The liquid has density ρ and electrical resistivity $\mathcal{R}$. The skin-friction coefficient in the generator is C_f.

The magnetic field induced by the generated current is assumed to be eliminated by compensating backstraps

which carry a return current I which, together with the shunt end currents I_s, cancels the current through the liquid metal.

2. Analysis

The emf induced between the electrodes at axial station x, where the width is a cm and the velocity is V cm/sec, is

$$E_0 = 10^{-8}aVB_0 \quad \text{volts} \tag{1}$$

where B_0 is in gauss.

Since the product aV is constant due to the incompressibility of the liquid metal, E_0 is also constant and is given by

$$E_0 = 10^{-8}a_1V_1B_0 \tag{2}$$

With E less than E_0, current flows in the direction of the induced emf and power is generated. Upstream and downstream of the generator, however, reverse currents flow since there is no field to induce an emf in those regions. As shown in Ref. 5, the actual generator with end currents is equivalent to a hypothetical generator with

Fig. 2. Idealized MHD generator

non-conducting fluid upstream and downstream and a shunt across the electrodes of resistance

$$R_s = \frac{\pi \mathscr{R}}{2b \ln 2} \qquad (3)$$

through which the current is

$$I_s = \frac{E}{R_s} \qquad (4)$$

Within the hypothetical generator bounded by non-conducting fluid, the current between the electrodes is

$$I + I_s = \int_0^L \frac{E_0 - E}{\mathscr{R}a} b \, dx \quad \text{amp} \qquad (5)$$

$$= \frac{E_0 b (1 - \mu)}{\mathscr{R}} \int_0^L \frac{dx}{a} \qquad (6)$$
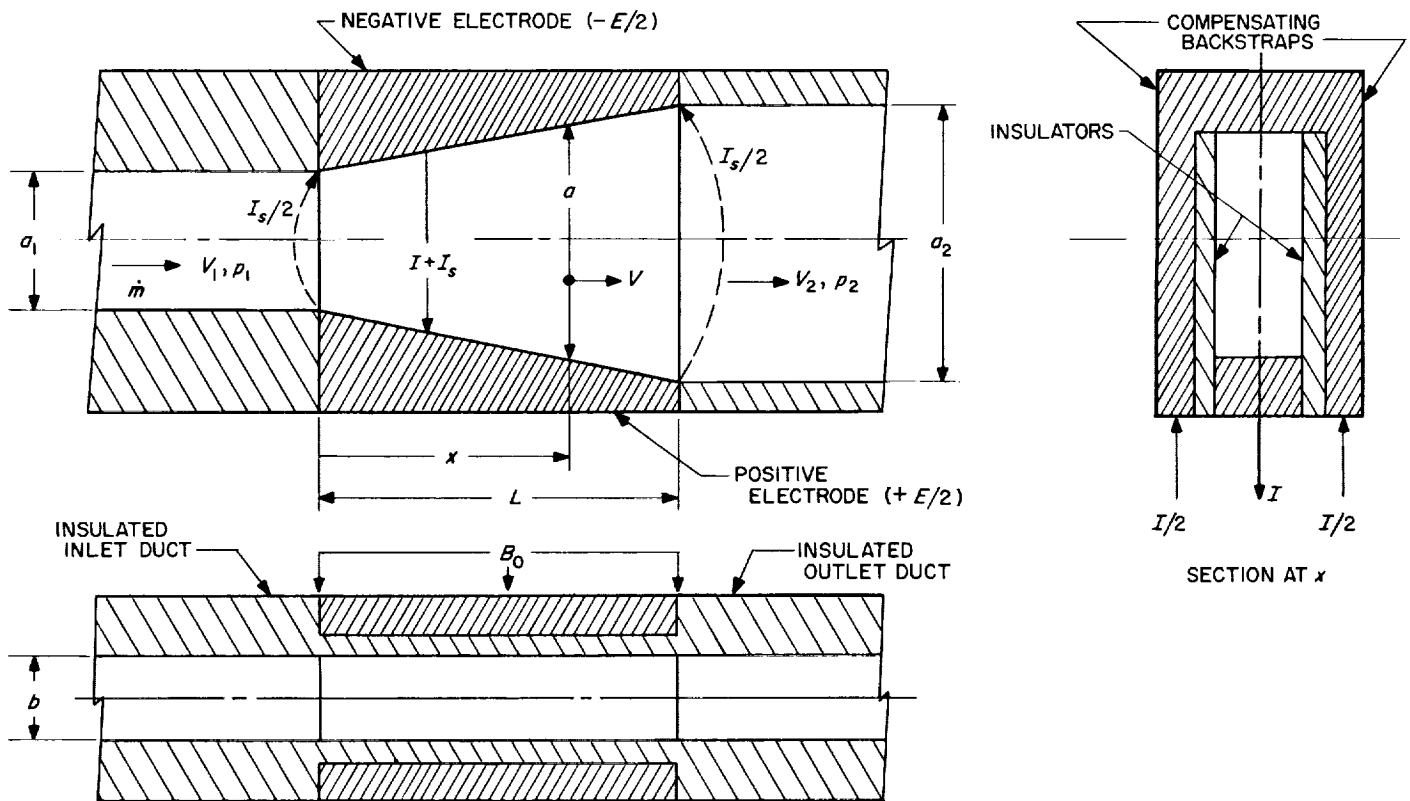
where μ is the loading E/E_0. For a linearly-tapered generator the integration yields

$$I + I_s = \frac{E_0 b L (1 - \mu)}{\mathscr{R}a_m} \qquad (7)$$

where a_m is the mean width given by

$$a_m = a_1 \frac{\dfrac{a_2}{a_1} - 1}{\ln \dfrac{a_2}{a_1}} \qquad (8)$$

The net electric power output of the generator is the product of the electrode voltage and the net current. Thus

$$P_e = EI \qquad (9)$$

Substituting Eqs. (2), (3), (4), and (7) into Eq. (9), the output power is

$$P_e = \frac{10^{-16}\mu(1-\mu)a_1^2 V_1^2 B_0^2 Lb}{\mathscr{R}a_m}\left[1 - \frac{2\mu a_m \ln 2}{\pi L(1-\mu)}\right] \qquad (10)$$

The power extracted from the fluid in the generator is equal to the gross electrical power $E_0 (I + I_s)$ plus the power dissipated in friction, P_f. Thus

$$P_0 = E_0 (I + I_s) + P_f \qquad (11)$$

The power extracted is evidenced by both velocity and pressure changes. Thus, P_0 is also given by

$$P_0 = \frac{\dot{m}}{2}(V_1^2 - V_2^2) + \frac{\dot{m}}{\rho}(p_1 - p_2) \qquad (12)$$

where m is the mass flow rate.

From Eqs. (9) and (11) the generator efficiency is

$$\eta = \frac{EI}{E_0(I+I_s)+P_f} \qquad (13)$$

The power dissipated in friction is the product of wall shear and fluid velocity integrated along the duct. Thus

$$P_f = 10^{-7} \int_0^L C_f \left(\frac{\rho V^2}{2}\right)(2a + 2b) V \, dx \quad \text{watts} \qquad (14)$$

with ρ in g/cm^3 and V in cm/sec.

Noting that $V = V_1 (a_1/a)$, the integration yields

$$P_f = 10^{-7} \rho V_1^3 C_f a_1^2 \frac{L}{a_2}\left[1 + \frac{b}{2}\left(\frac{a_1 + a_2}{a_1 a_2}\right)\right] \qquad (15)$$

Substituting Eqs. (2), (3), (4), (7), and (15) into Eq. (13), the theoretical efficiency is

$$\eta = \mu \, \frac{1 - \dfrac{2\mu a_m \ln 2}{\pi L (1-\mu)}}{1 + \dfrac{10^9 \rho V_1 C_f \mathscr{R} a_m}{B_0^2 (1 - \mu) a_2 b}\left[1 + \dfrac{b}{2}\left(\dfrac{a_1 + a_2}{a_1 a_2}\right)\right]} \qquad (16)$$

3. Experimental Generator

Two generators have been built for testing with single-phase NaK, one with a straight channel for operation at constant velocity and the other with a diverging channel for operation at constant pressure. Only the straight generator has been tested so far.

Fig. 3 is a photograph of the straight generator bolted to its inlet and outlet ducts, and Fig. 4 shows the disassembled generator. The center part of the generator is a copper block grooved to form a channel bounded by the negative electrode and the backstraps. The end flanges bolt to the ducts and are sealed with O-rings. All internal surfaces of the copper block, except the electrode face, are coated with polyurethane insulation. The ducts neck down to the generator cross section about two inches from the ends of the electrodes and are also coated inside.
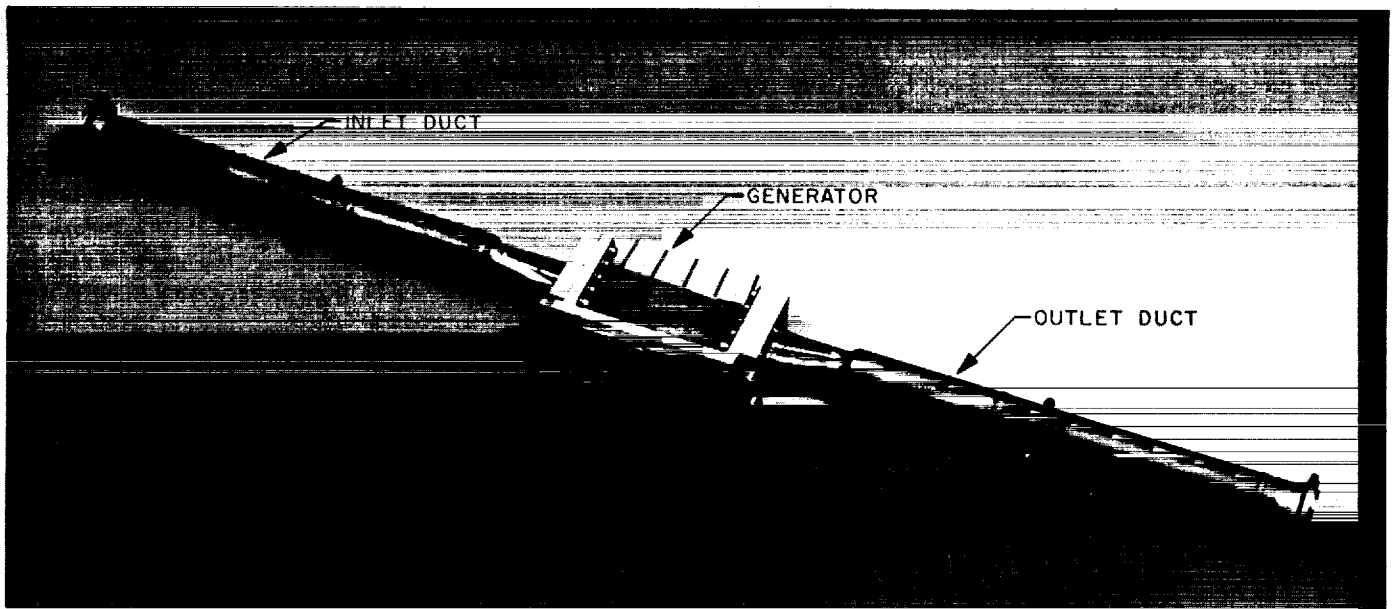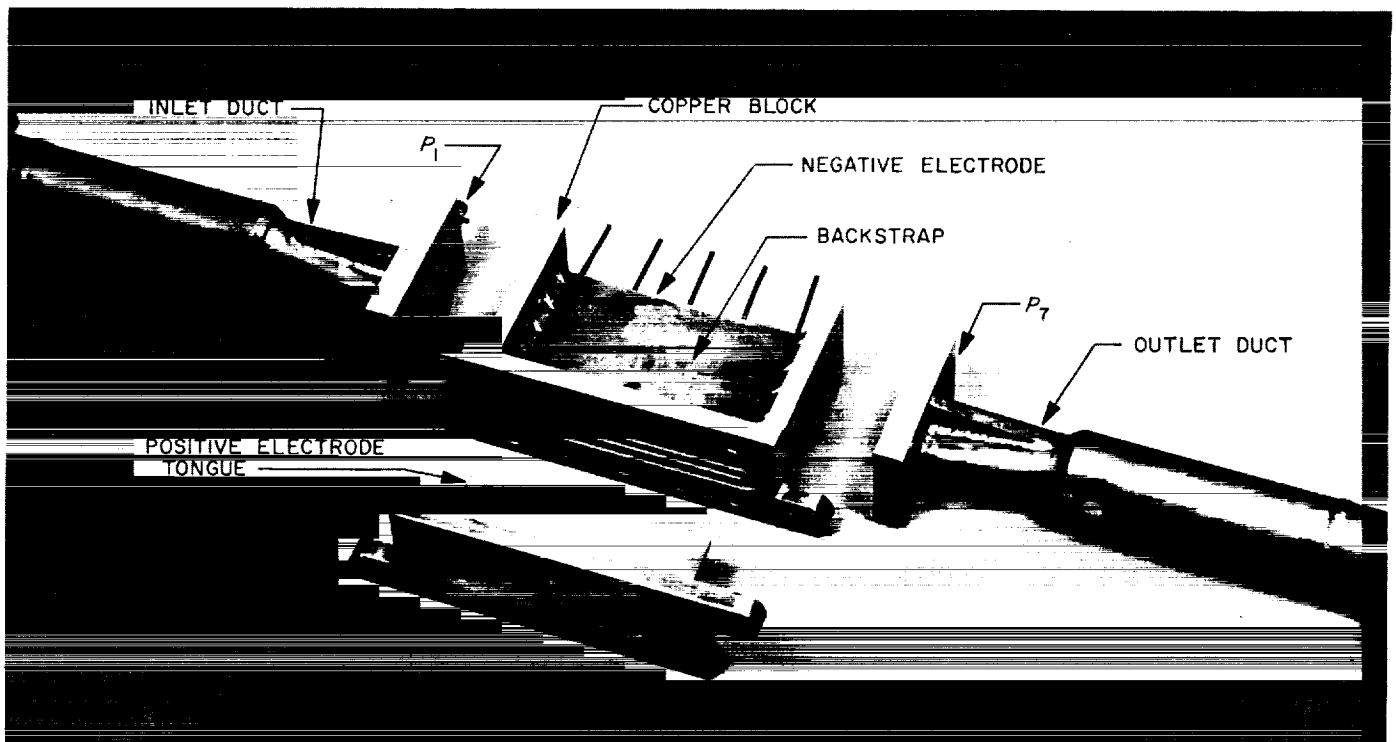
Fig. 3. Experimental generator

Fig. 4. Disassembled generator

The positive electrode is a tongue which slides into a slot in the generator block and is sealed by an O-ring. The tongue and its mounting flange are insulated from the center block by the polyurethane coating which extends through the slot and beyond the O-ring groove.

The current path in the generator is from terminals on each end of the center block, through the backstraps, into the negative electrode, through the NaK, into the tongue, and to terminals on each end of the tongue flange.

Seven pressure taps are provided, one in each duct flange and five in the center block.

The dimensions of the constant velocity generator are

$$a_1 = a_2 = a_m = 1.575 \text{ in.}$$

$$b = 0.200 \text{ in.}$$

$$L = 5.854 \text{ in.}$$

The latter dimension is the length of the electrodes. The magnet poles, however, are 6.0 in. long, and the half-field points are 7 in. apart.

For this generator the loading which gives maximum efficiency is approximately $\mu = 0.7$. For that loading, for the above dimensions, and for a NaK density of 0.877 g/cm^3 (hydrometer measurement on the NaK used in the experiments) and NaK resistivity of 34 μohm–cm (Ref. 6), the theoretical output power according to Eq. (10) is

$$P_e = 0.0125 \, V_1^2 \, B_0^2 \quad \text{watts} \tag{17}$$

where V is in ft/sec and B_0 is in kilogauss.

From Eq. (16) the theoretical efficiency is

$$\eta = \frac{0.506}{1 + 6.72 \dfrac{V_1 C_f}{B_0^2}} \tag{18}$$

The experimental output power is the product of the output current and voltage, the latter corrected to the value at the electrode interface with the NaK. The experimental input power is calculated from the flow rate and pressure drop using Eq. (12) which, for the constant-velocity generator, reduces to

$$P_0 = 3.564 \, \dot{m} \, (p_1 - p_7) \quad \text{watts} \tag{19}$$

where $\dot{m}$ is in lb/sec and p_1 and p_7 are the pressures, in psi, at the inlet and outlet pipe flanges, respectively.

4. Test Setup

Fig. 5 is a photograph of the experimental generator installed in the NaK test facility. NaK at pressures up to 1000 psi and at flow rates up to 40 lb/sec is supplied from a nitrogen-pressurized tank in an adjoining cubicle to the generator inlet duct in the foreground. The flow is started and stopped by a 1½-in. bellows-sealed, pneumatically actuated valve. Upstream of the valve a turbine meter measures the NaK flow rate to an accuracy of ± 0.5%.

The generator is clamped between the poles of an electromagnet with 6- × 1.7-in. pole faces supplying fields up to 8500 gauss with a maximum variation of ± 8% throughout the generator channel. The generator output power is dissipated across a pair of load resistors consisting of water-cooled copper tubes. The current in each resistor is determined from the voltage drop across it. These voltage drops and the output voltage of the generator are recorded to an accuracy of ± 0.5% on a digital data system in the control room. Pressures are measured by strain gage transducers and also recorded on the digital system.

Downstream of the generator a remotely-actuated throttling valve controls the pressure of the NaK returning to a receiver tank in the adjoining cubicle. Available run durations are determined by the 1400 lb of NaK which can be transferred per run. After transferring that amount, the supply tank is vented and refilled from the receiver tank. Fig. 6 shows the control room for the generator tests. At the right is the console for controlling the pressurizing and venting of the NaK tanks and for controlling the NaK flow. Beyond the console is the magnet power supply. On the left is the digital data recording system.

5. Generator Tests

The tanks were loaded with 1700 lb of NaK on March 3 and initial generator runs were made on March 6. On March 9 twelve runs were made with a nominal 50-μohm load resistance (theoretical $\mu = 0.67$) at various velocities and fields, and accurate data was obtained for all parameters except the sixth pressure tap which was plugged. Following these runs a 95-μohm load-resistor pair ($\mu = 0.76$) was installed, but after two runs the output voltage dropped to a value indicating an internal short circuit in the generator, and testing was terminated. Subsequent disassembly of the generator showed that the polyurethane coating had failed near one corner of the
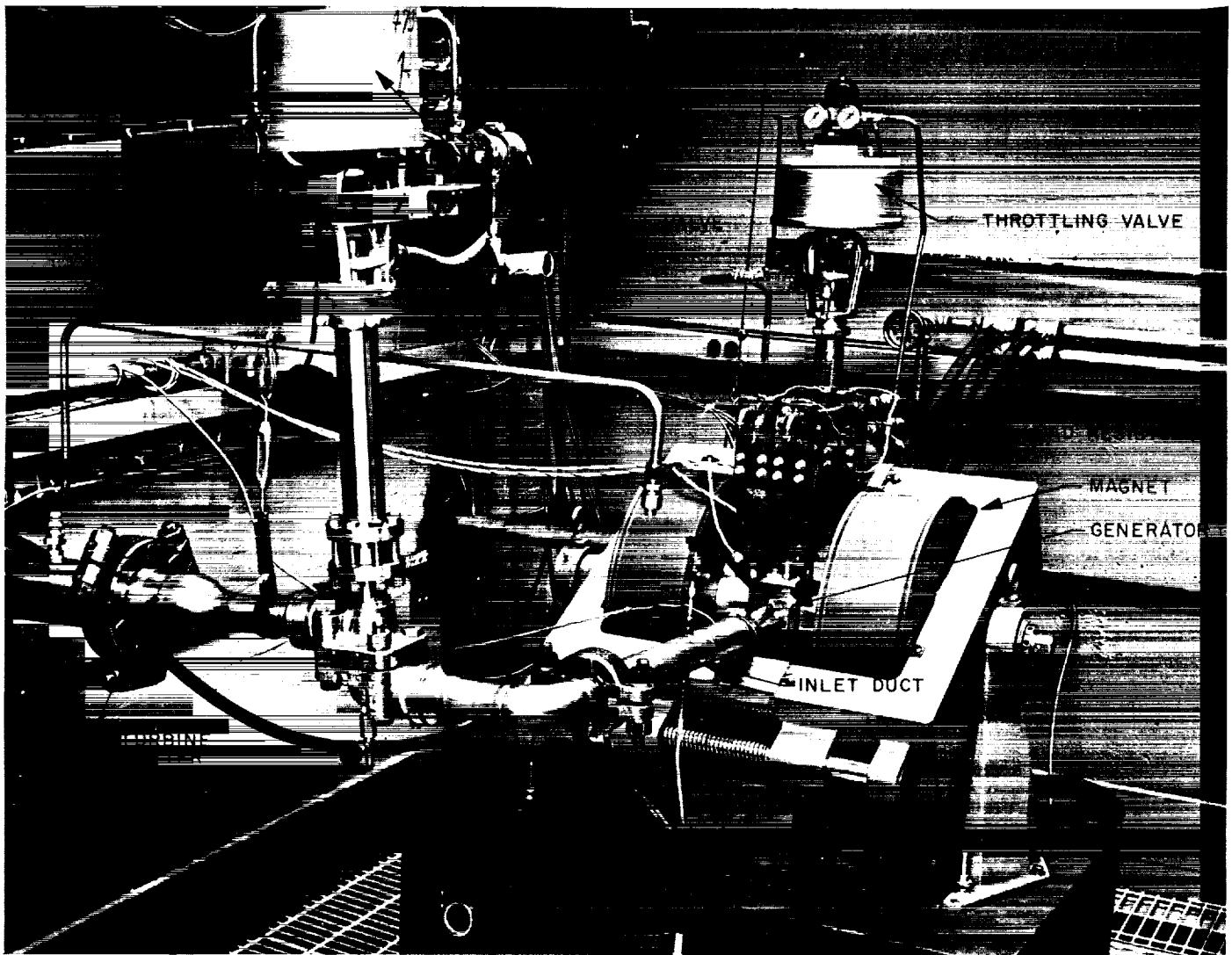
Fig. 5. Generator test arrangement

tongue, possibly due to insufficient clearance and result-ing damage on assembly. Greater clearances will be employed in the future.

The performance on the twelve runs with the 50-μohm load resistance was excellent. Power outputs from 200 to 2400 watts were obtained, and the efficiency at the lowest power was 49.5%, only one percentage point below the theoretical maximum predicted by Eq. (18). The output voltage at 2400 watts was 0.36 volts.

The output power and efficiency with the 50-μohm load resistor were measured at field strengths of 2250, 3580, 5080, and 8500 gauss at several different velocities obtained by varying the NaK supply tank pressure. The

generator inlet pressure was limited to 150 psi for these runs. The measured loading μ varied from 0.62 to 0.72.

Fig. 7 compares the measured output powers with the theoretical values calculated from Eq. (17) for the fields used in the tests. The agreement is within 10% for most of the data points.

Fig. 8 compares the measured efficiencies with the theoretical values calculated from Eq. (18) for $C_f = 0.004$. This skin-friction coefficient corresponds to fully devel-oped pipe flow at the velocities of interest and was confirmed by prior water tests. It is seen that although the measured efficiencies are within six percentage points
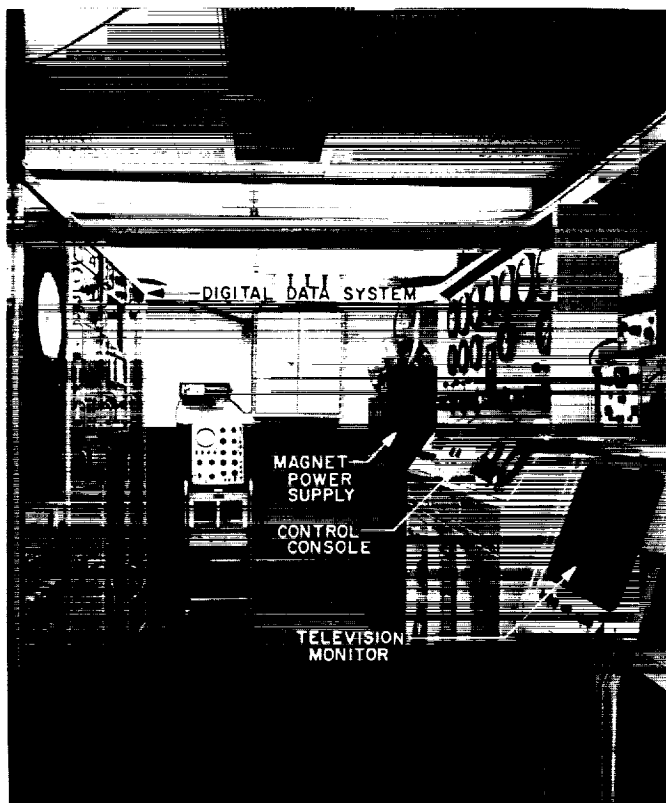
Fig. 6. Control room for generator tests

Fig. 7. Comparison between experimental and theoretical output power of constant-velocity generator

of the theoretical values they tend to drop off more rapidly with velocity.

Close agreement of the theoretical and experimental efficiencies could be obtained by increasing C_f to 0.006 in. in Eq. (18), but the real reason for the discrepancy appears to be incomplete compensation of the induced field. For the high and low efficiency points, A and B respectively in Fig. 8, Fig. 9 presents the corresponding pressure variation along the generator channel. It is seen that for the high efficiency point the pressure drop is linear, indicating uniform power generation along the channel. However, for the low efficiency point B almost all of the pressure drop, hence power generation, occurs at the downstream end of the generator. Analysis of this run showed that there was an induced field varying from 1500 gauss opposite the applied field at the upstream end to the same amount aiding the applied field at the downstream end. The increased I^2R loss due to the nonuniform power generation was calculated and found to account for the entire difference between the measured and theoretical efficiency for this run. Thus, it can be concluded that the skin-friction coefficient was not increased by MHD effects.

The source of the incomplete compensation is believed to be the end flanges on the generator which drew axial currents from the backstraps.

Another source of loss not considered in the analysis is the shunting effect of the boundary layer. For a fully developed one-seventh-power velocity profile, however, the efficiency at $\mu = 0.7$ would be decreased only four percentage points, too small an effect to be conclusively observed in these experiments.

The main conclusions from the tests of the constant-velocity generator are:

(1) The performance is correctly predicted by the analysis of Ref. 5, provided care is taken to fully compensate the induced field.

(2) There is no increase in the skin-friction coefficient due to the operation of the generator.
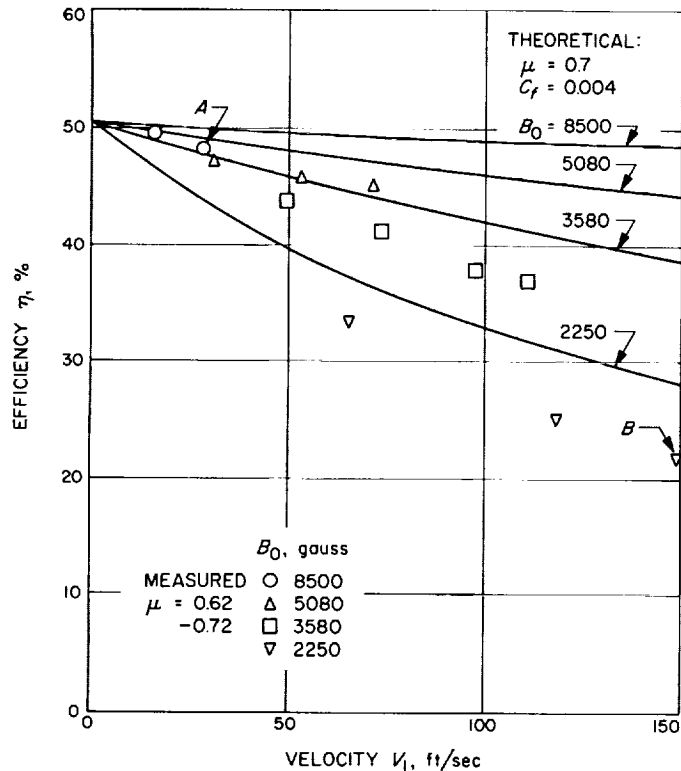
Fig. 8. Comparison between experimental and theoretical efficiency of constant-velocity generator
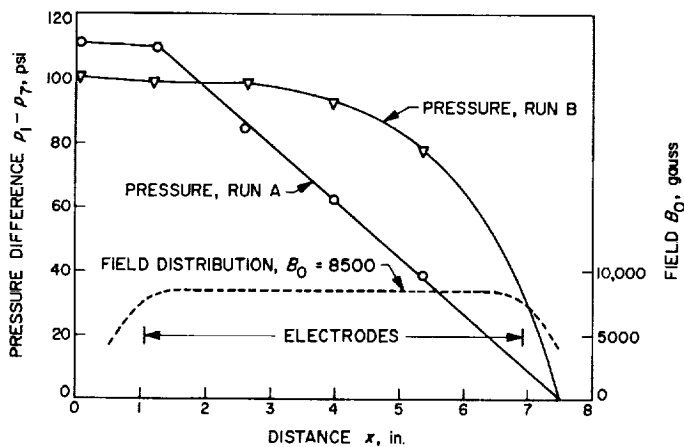
Fig. 9. Measured pressure profile in constant-velocity generator

B. Lithium-Boiling Potassium Loop

N. E. Kogen

The columbium-1% zirconium (Cb-1% Zr) lithium-boiling potassium loop described in previous SPS issues is being modified to accept a potassium-vapor-driven turboalternator which will be supplied by Aeronutronics Division of Philco Corporation under contract BF4-217030.

The turbine is designed to operate under the following conditions:

Working fluid	Potassium vapor
Turbine inlet temperature, °F	1900
Turbine exhaust temperature, °F	1500
Turbine weight flow, lb/sec	0.0324
Thermal power to potassium, kw	30
Operating life, hr	10,000

The materials of construction are Cb-1% Zr and molybdenum for all parts wetted by the potassium vapor. Liquid metal lubricated Kennametal K-94 bearings will be used. Bearing leakage flow is discharged into the turbine exhaust. Auxiliary cooling will be supplied to the alternator stator by cooling oil at a maximum temperature of 300°F.

The turbine wheel is 6.3 in. in diameter and consists of a solid molybdenum disc with 120 impulse buckets cut into the rim. The turbine is a partial admission two-stage re-entry turbine having two nozzle sections and a single rotor. The design speed of the turbine is 12,000 rpm, which results in a safety factor of 10 with a 10,000-hr rupture stress of 40,000 psi at a blade temperature of 1750°F.

The alternator is to be a three-phase 200-cycle 240/416-v induction alternator rated at 1.0 kw and excited by static capacitors. The rotor diameter will be 2.875 in. and the length 1.875 in.

At the design speed of the turboalternator the turbulent bearing effect will be small based on bearing analysis and extrapolation of testing data conducted at 8000 rpm.

The turbine efficiency is calculated at 25%, including losses due to wetness of the vapor. The estimated electrical efficiency of the alternator is 70% and the electrical
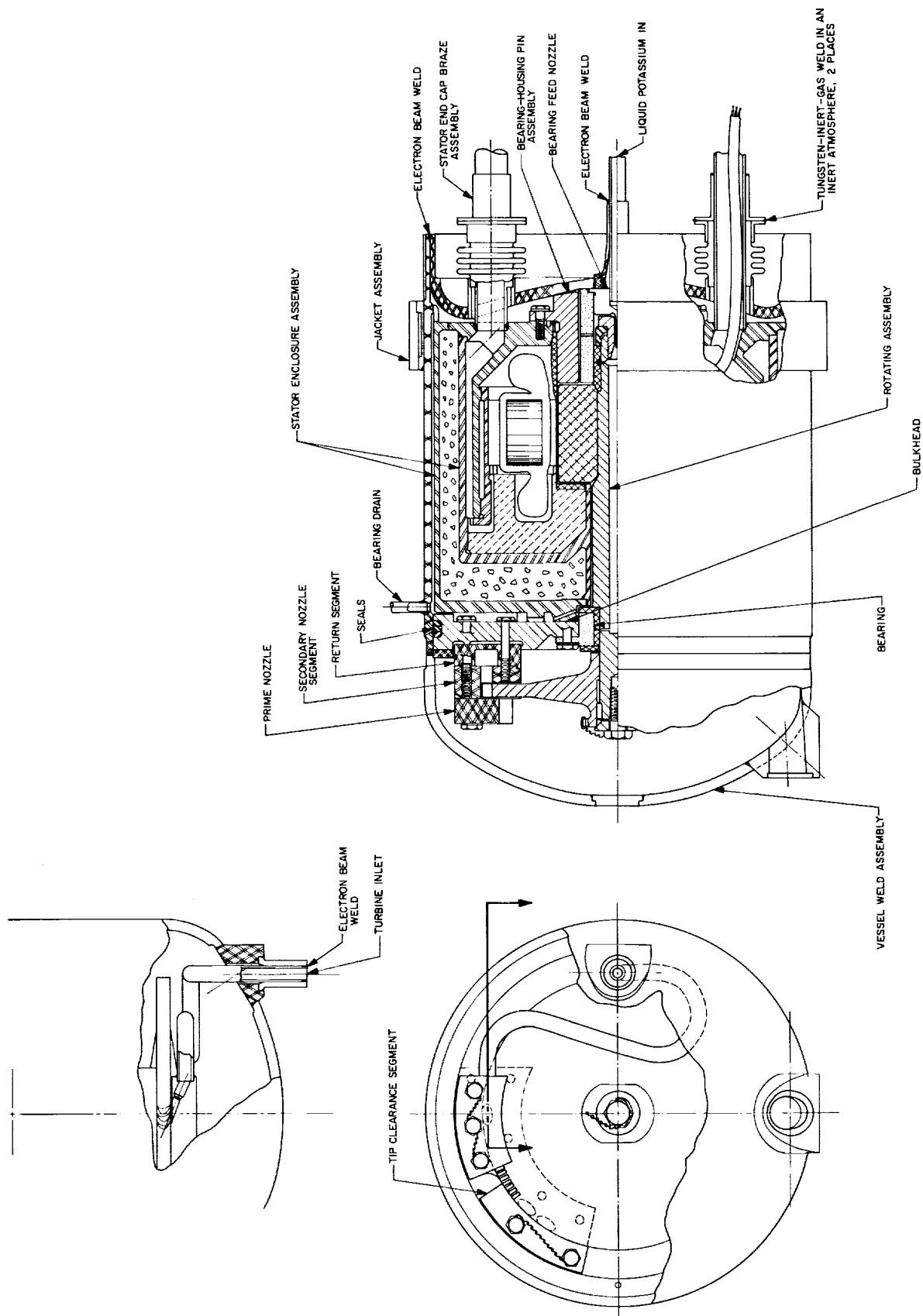
Fig. 10. Potassium vapor turboalternator

output of the alternator is approximately 1 kw at the given flowrate of 0.0324 lb/sec. With a slight modification of the turbine and an increased flowrate to 0.09 lb/sec, an electrical output of 3 kw can be expected.

The turboalternator has a growth potential to approximately 25 kw with full admission operation. The bearing, alternator, and turbine technology will be applicable to larger turboalternator units. In addition, answers obtained with regard to turbine blade erosion, alternator sealing methods, and liquid metal bearing operation are immediately applicable to the design of larger power units.

Fig. 10 is a drawing of the turboalternator. Final drawings have been submitted and approved and fabrication has started. The turboalternator will be assembled first with roller bearings in place of the Kennametal and operated with nitrogen gas before delivery.

The installation in the loop will include a bypass line around the turboalternator to allow for testing the loop both with and without the turboalternator operating.

References

1. Elliott, D., and Cerini, D., "Liquid MHD Power Conversion," *SPS 37-17*, Vol. IV, pp. 199–203, Jet Propulsion Laboratory, Pasadena, California, October 30, 1962.

2. Elliott, D., Cerini, D., McNary, C., Stephenson, R., and Weinberg, E., "Liquid MHD Power Conversion," *SPS 37-21*, Vol. IV, pp. 114–121, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.

3. Elliott, D., Cerini, D., Otte, H., and Weinberg, E., "Liquid MHD Power Conversion," *SPS 37-23*, Vol. IV, pp. 132–135, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

4. Elliott, D., Cerini, D., Hays, L., and Weinberg, E., "Liquid MHD Power Conversion," *SPS 37-25*, Vol. IV, pp. 106–111, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

5. Sutton, G. W., Hurwitz, H., and Poritsky, H., *Electrical and Pressure Losses in a Magnetohydrodynamic Channel Due to End Current Loops*, TIS Report R59SD431, General Electric Co., Philadelphia, Pa., July 22, 1959.

6. Drugas, P. G., Rehn, I. R., and Wilkinson, W. D., *Resistivity of NaK*, Report No. ANL-5115, Argonne National Laboratory, Argonne, Illinois, October 2, 1953.

SPACE SCIENCES DIVISION

XII. Space Instruments Systems

A. Development of a Hysteresis Motor for Spacecraft Tape-Recorder Applications

E. Bahm

1. Requirements

A hysteresis motor has been developed for use in spacecraft tape recorders. Data for the recorders are given in Table 1.

The motor must be: (1) capable of operating at two different speeds for recording and playback of the recorder, starting as fast as possible at the low speed during playback; and (2) designed for smooth operation, although some speed variations during synchronous operation were tolerable. Maximum reliability and highest possible efficiency were required.

Table 1. Motor requirements

Unit	Measurement
Load of the tape recorder	4 g cm at both speeds
Speed for record mode	8000 rpm
Speed for playback mode	1000 rpm
Power supply Frequency	400 and 50 cps, two-phase, square-wave
Available motor input power	2.5 w

2. Reliability of a Motor Operating in a Space Environment

After a motor is properly designed, carefully manufactured, and tested, the performance of the bearings is the only major obstacle to the achievement of a high

order of reliability. Therefore, the bearing problem needs the most careful evaluation. Lifetime tests with bearings and motors in a vacuum of 2×10^{-6} torr have been performed in our laboratories using various recommended lubricants. As a result of these and other tests, it was felt that double-shielded, over-dimensional ball bearings heavily lubricated with General Electric Versilube F50 should give the best results. Versilube F50 has a very low vapor pressure and showed a lifetime of 1.5 yr in a vacuum of 10^{-8} torr. Heavy lubrication is considered necessary if the motor is subjected to sterilization prior to launch, since some lubricant is lost during the high-temperature application.

The bearings are not preloaded, but are carefully shimmed to establish a very small axial play. Selected bearings with minimum radial play have been used. In spite of their relatively large size, these bearings develop very little friction torque and therefore operate very coolly. Because of the light weight of the rotor and the large bearings, a vibration test produced no effect on the bearing performance.

Testing under vacuum conditions revealed a continuous increase of friction torque during operation of the bearings. This increased torque must be considered normal and does not necessarily indicate an approaching failure of the bearings. However, it must be allowed for in the design of the motor. The load of the tape recorder is due to friction load and will increase continuously during a long voyage in space. Therefore, the
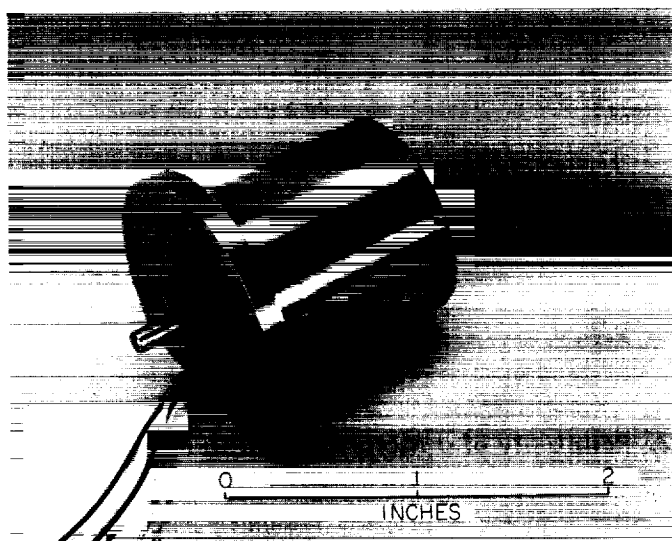
design of the motor must be such that a friction load of at least three times the tape-recorder load as measured in the laboratory can be tolerated.

3. Final Design of the Hysteresis Motor

Following a design study, an engineering model of the motor was built and tested, and a design review for the final unit was conducted. This motor has been designed to meet the specifications for the *Mariner C* (Mars mission, 1964) flight equipment. Data for the unit (Fig. 1) are given in Table 2.

Figs. 2 and 3 show the efficiency and pull-in torque versus motor voltage at 400 and 50 cps, respectively.

4. Cogging of the Motor at Low Speed

The amount of speed variation of the motor during synchronous operation is called cogging C, and is defined as follows:

$$C_{rms} = 2^{(1/2)} \frac{n_{max} - n_{min}}{n_{max} + n_{min}} 100\%$$

Table 2. Design specifications

Unit	Measurement
Outside diameter, in.	1.06
Length (not including shaft), in.	1.25
Shaft diameter, in.	0.123
Shaft extension beyond motor face, in.	0.375
Number of poles	6
Maximum efficiency at 27 v, 400 cps, square wave, pull-out torque, and 29°C, %	55
Maximum efficiency at 6 v, 50 cps, square wave, pull-out torque, and 29°C, %	8.4
Operating voltage at 400 cps, v	30
Operating voltage at 50 cps, v	8.5
Efficiency at operating voltage, 400 cps, pull-out torque, and 29°C, %	53
Efficiency at operating voltage, 50 cps, pull-out torque, and 29°C, %	5.2
Input power at operating voltage, 400 cps, and pull-out torque, w	2.1
Start time with no load at 8.5 v and 50 cps, msec	22

Fig. 1. Hysteresis motor
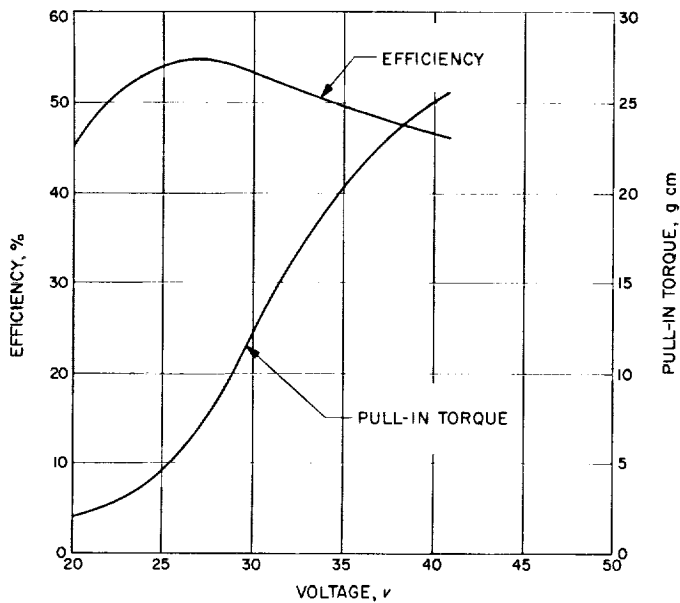
Fig. 2. Efficiency and pull-in torque versus phase
voltage with a 400-cps square-wave input
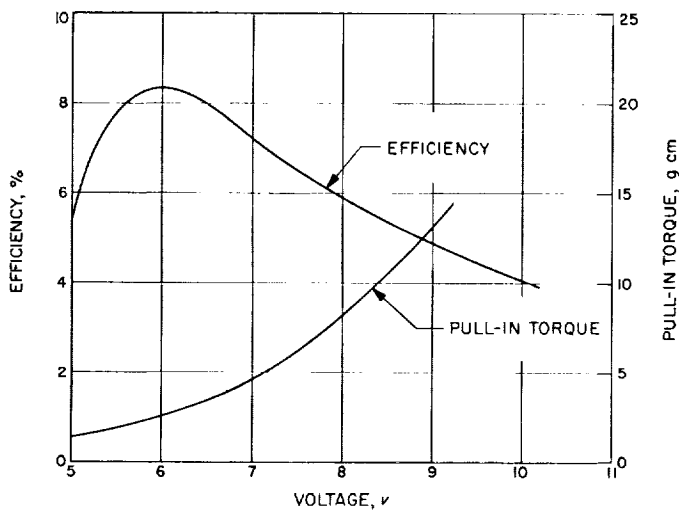
Fig. 3. Efficiency and pull-in torque versus phase
voltage with a 50-cps square-wave input

where n_{max} and n_{min} are the maximum and minimum
motor speeds, respectively.

Cogging is a typical phenomenon of small motors
running at low speeds. In many cases it is the main
source of flutter in a recorder. A motor operating from
a square-wave power supply shows more cogging
than one operating from a sine-wave supply because the
voltage does not vary smoothly in the former. Cogging
is measured by recording pulses on magnetic tape, which

UPPER TRACE: I-msec/cm OUTPUT; 100-cps,
10.5-v, SQUARE-WAVE INPUT
LOWER TRACE; 2-msec/cm OUTPUT; 50-cps,
8.5-v, SQUARE-WAVE INPUT
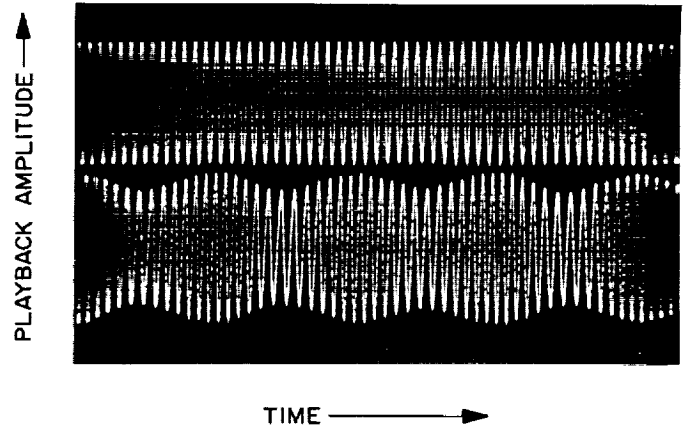
Fig. 4. Oscillogram of a playback signal of pulses
on magnetic tape

is then played back with the test motor driving the re-
corder at test frequency. With a perfectly smooth-running
motor the playback signal is a true sine wave within the
limits of other sources of flutter in the recorder and the
uniformity of the magnetic tape coating. These sources
of error, however, can be eliminated during evaluation
of the oscillograms, since they show either a different
periodicity or none at all. An oscillogram of the play-
back signals is shown in Fig. 4.

Cogging is dependent on the drive-frequency and
voltage. Measurement versus frequency is given in Fig. 5.
Because of the damping effects of the motor reactance
and the rotor inertia, significant cogging occurs only at
low drive frequencies. From the upper trace in Fig. 4, it
can be seen that the operation of the motor is already
perfectly smooth at 100 cps, since no periodic pulsation
of the amplitude can be observed. Cogging increases
rapidly with the drive-voltage, as can be seen from
measurements in Fig. 6. This increase is due to increased
field intensity in the motor.

5. Cogging of Extremely Small Motors

At 50 cps, C_{rms} of the developed motor is 38% at
operating voltage, which may be tolerable for a digital
data recorder. However, a vibration study revealed that
the cogging problem becomes severe for very small tape
recorders which have much smaller motors than that
described and which are running at low speed. For such
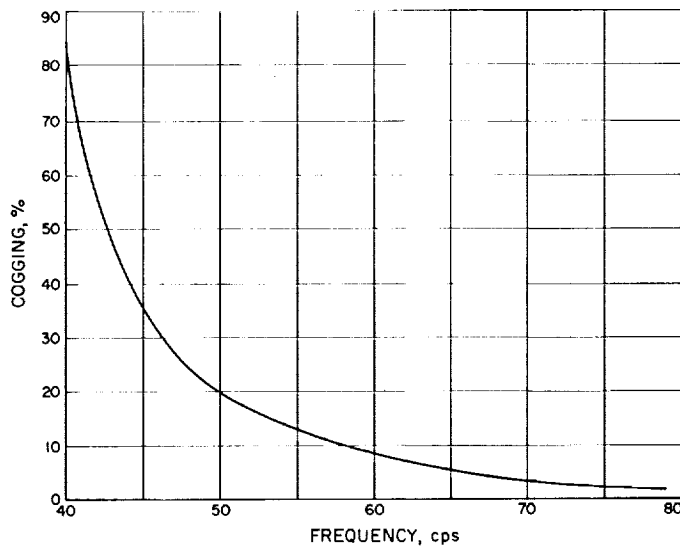
Fig. 5. Cogging of the hysteresis motor versus drive frequency at constant field intensity
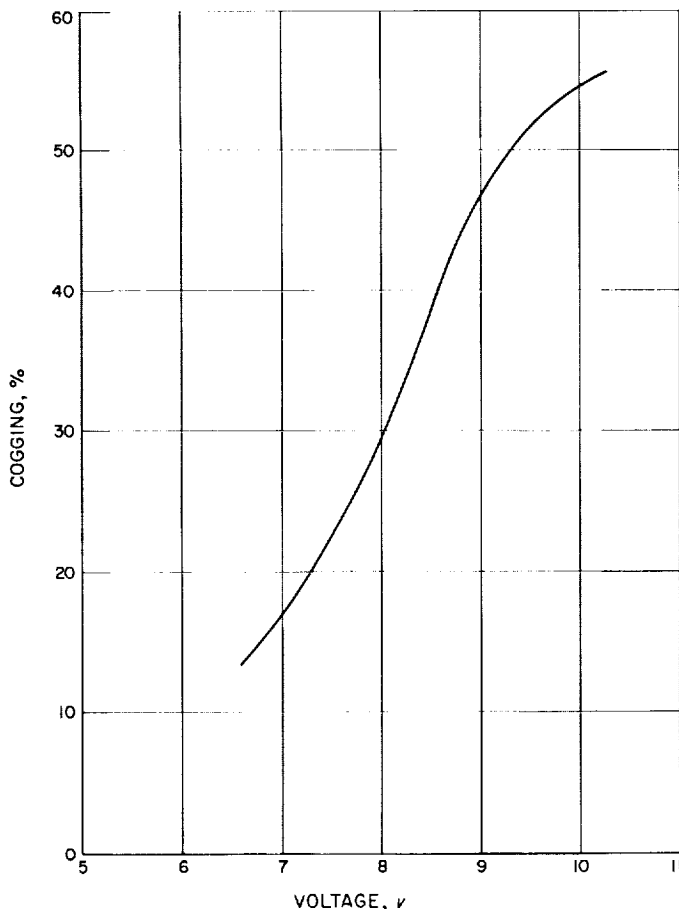
Fig. 6. Cogging of the hysteresis motor versus phase voltage at 50 cps

applications the dc motor is more suitable. It is felt that the major breakthrough in the development of the brushless dc motor has been accomplished, and this motor is expected to be superior to the hysteresis motor for certain tape-recorder applications.

6. Magnetic Field Outside of Motor

As the motor has a permanent magnetic rotor, a static magnetic field is developed in the neighborhood, if no power is applied. Maximum field intensity at 3 ft was found to be 2 γ, which is tolerable. During operation the motor does not produce any frequency below the drive frequency.

B. Fiber Optics

R. Y. Wong

1. Introduction

This summary describes a portion of the work on the theoretical analysis, design and development of a fiber optic system for space application. Although the mechanism of light transmission through an optical dielectric cylinder has been well known for some time, only in recent years has industry developed the advanced techniques necessary to produce fiber optics capable of withstanding the varying environments encountered in space applications. A fine fiber, when surrounded by a coating of dielectric material of lower index of refraction, transmits light energy by the mechanism of total internal reflection; i.e., light rays propagate through the fiber by a series of reflections from wall to wall. The fibers, when fused together to form a dielectric bundle, can convey an image from one place to another. The bundle is capable of gathering a large quantity of light energy since each fiber is, in essence, serving as an energy collector.

On the basis of thermodynamical reasoning, no increase in photometric efficiency is achieved as light is transmitted through the fibers. In fact, since various losses occur as the light is conducting along the fiber, the factors affecting transmission efficiency are important in the study of fiber optics.

Fig. 7. Cylindrical fiber

$$\gamma_{\perp} = \left\{ \frac{\left(\frac{N_1}{N_0}\right)^2 \cos\theta - \left[\left(\frac{N_1}{N_0}\right)^2 - \sin^2\theta\right]^{1/2}}{\left(\frac{N_1}{N_0}\right)^2 \cos\theta + \left[\left(\frac{N_1}{N_0}\right)^2 - \sin^2\theta\right]^{1/2}} \right\}^2 \qquad (3)$$

$$\gamma_{||} = \left\{ \frac{\cos\theta - \left[\left(\frac{N_1}{N_0}\right)^2 - \sin^2\theta\right]^{1/2}}{\cos\theta + \left[\left(\frac{N_1}{N_0}\right)^2 - \sin^2\theta\right]^{1/2}} \right\}^2 \qquad (4)$$
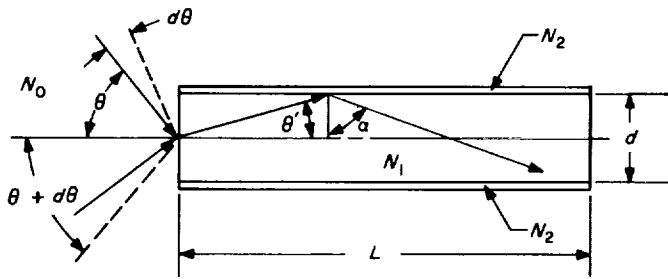
2. Transmission Characteristics

Fig. 7 shows a cylindrical fiber of diameter d and length L. Such a fiber is assumed to be representative of the many fibers fused together to form a cylindrical bundle. A cone of light having a half-cone angle of θ strikes the fiber surface and a portion of the incident ray passes through the surface at a refraction angle of θ'. As the ray is traveling through the fiber, it bounces from wall to wall and eventually exits through the end surface. The various losses as the ray is conducting through the fiber are to be analyzed on the basis of meridional ray consideration.

Light conducts through the bundle most effectively when the light at the entrance surface is incident at a half-cone angle of equal to or less than the critical angle θ_c. Rays incident at greater angles will be mostly reflected outside the fiber. The critical angle is given by

$$\theta_c = \sin^{-1} \left[\frac{(N_1^2 - N_2^2)^{1/2}}{N_0} \right] \qquad (1)$$

where N_1, N_2 and N_0 are, respectively, the indices of refraction of the core fiber, coating, and medium surrounding the bundle.

As the ray is entering and leaving the fiber, energy is lost due to dielectric boundary or surface reflection. This loss is a function of the indices of refraction of N_0 and N_1. If the fiber surface is smooth, the fractional loss due to surface reflection at the entrance is

$$R(\theta) = \frac{1}{2} [\gamma_{\perp} + \gamma_{||}] \qquad (2)$$

where $\gamma_{\perp}$ and $\gamma_{||}$ are the Fresnel coefficients of refraction for plane-polarized incident radiation when the E-vector of the electromagnetic wave is, respectively, perpendicular and parallel to the plane of incidence.

$R(\theta)$ of Eq. (2) is the Fresnel coefficient of reflection for unpolarized incident radiation. The fraction of the energy refracted through the entrance surface is therefore $[1 - R(\theta)]$. The fraction of the energy leaving the exit surface after a second refraction is proportional to $[1 - R(\theta)]^2$. Numerical evaluations of $[1 - R(\theta)]^2$ indicated that the transmission efficiency is fairly constant for $\theta \le 50$ deg and decreases very rapidly for $\theta > 50$ deg. Light energy therefore, propagates through fiber optics bundles most efficiently when it strikes the entrance surface at an incidence half-cone angle of equal to or less than 50 deg.

As the ray is propagating through the fiber by a series of internal reflections, the length of the ray path $P(\theta)$ through a fiber having a length L is

$$P(\theta) = \frac{L}{\left[1 - \left(\frac{\sin\theta}{N_1}\right)^2\right]^{1/2}} \qquad (5)$$

Energy is lost along the path due to glass absorption. As the length of the ray path increases, the fractional transmitted energy τ_A decreases exponentially

$$\tau_A = \exp \left\{ \frac{-AL}{\left[1 - \left(\frac{\sin\theta}{N_1}\right)^2\right]^{1/2}} \right\} \qquad (6)$$

where A is the coefficient of absorption of the glass. Since the transmission efficiency drops off very rapidly as AL increases, it is desirable to fabricate long fibers from glass of low absorption to minimize this loss.

In spite of a well-defined interface between the core and coating, energy is lost due to absorption by the coating each time the ray is reflected. For a ray with an incident angle of θ, the number of internal reflections

before the ray reaches the exit end can be calculated by the following equation:

$$N(\theta) = \frac{L}{d} \tan\left[\sin^{-1}\left(\frac{N_0}{N_1}\sin\theta\right)\right] \qquad (7)$$

The fractional energy transmitted after $N(\theta)$ number of reflections is $[R_1(\theta)]^{N(\theta)}$. Here, $R_1(\theta)$ is the Fresnel coefficient of reflection for unpolarized light as evaluated by Eqs. (2), (3), and (4) with N_2/N_1 substituted for N_1/N_0 and α substituted for θ, where

$$\alpha = \frac{\pi}{2} - \theta'$$

$$\theta' = \sin^{-1}\left(\frac{N_0}{N_1}\sin\theta\right)$$

$$\alpha = \frac{\pi}{2} - \sin^{-1}\left(\frac{N_0}{N_1}\sin\theta\right)$$

This loss does not include the effect of the multipath reflections between interfaces of the core and coating and the neighboring core and coating. It will be shown in a future analysis that the magnitude of these higher order reflections is generally small and can be neglected.

To derive a transmission efficiency equation with the losses taken into consideration, it is convenient to consider the incident energy contained in a hollow cone between θ and $d\theta$ as shown on Fig. 7. The energy contained in this cone is equal to the solid angle times the energy per steradian or

$$dE(\theta) = 2\pi I(\theta) \sin\theta \, d\theta$$

where $I(\theta)$ is the energy distribution of the incident rays as a function of the incident angle (θ).

$$\tau = \frac{E(\theta)_{OUT}}{E(\theta)_{IN}} = \frac{\displaystyle\int_0^{\theta_0} 2\pi I(\theta) \tau_1 \sin\theta \, d\theta}{\displaystyle\int_0^{\theta_0} 2\pi I(\theta) \sin\theta \, d\theta}$$

Fig. 8. Transmission efficiency as a function of half-angle of incident light cone for $d = 10\ \mu$

where τ is the total transmission efficiency, τ_1 is the composite transmission efficiency due to the various losses and θ_0 is the half-cone angle of the incident light.

To study the effects of the various parameters on the total transmission efficiency, Eq. (8) was solved by the IBM 7094 Computer using numerical integration for a number of cases. In the computation, the incident flux was assumed to be uniform; $I(\theta) = $ constant, A was chosen to be 1% per inch and N_0 was set to 1.00 (for vacuum operation). The results are presented as curves on Figs. 8–12. Fig. 8 shows the transmission efficiency as a function of the incidence half-cone angle for fiber bundles of 2, 4, and 8 in. in length and fiber size of 10 μ.

Figs. 9 and 10 show the transmission efficiency as a function of the incidence half-cone angle with the same

$$= \frac{\displaystyle\int_0^{\theta_0} I(\theta)\left[1 - R(\theta)\right]^2 \left[R_1(\theta)\right]^{N(\theta)} \exp\left\{\frac{-AL}{\left[1 - \left(\dfrac{\sin\theta}{N_1}\right)^2\right]^{1/2}}\right\} \sin\theta \, d\theta}{\displaystyle\int_0^{\theta_0} I(\theta) \sin\theta \, d\theta} \qquad (8)$$
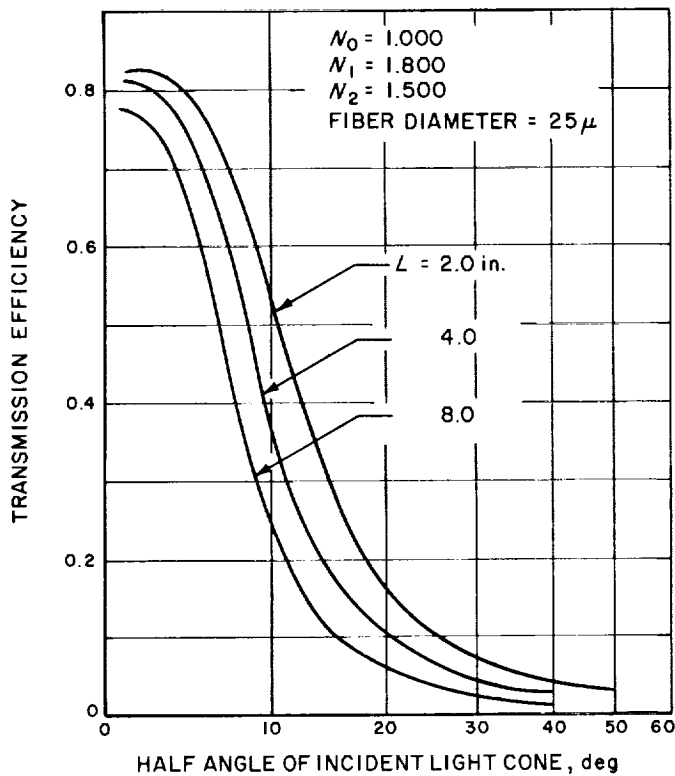
Fig. 9. Transmission efficiency as a function of half-angle of incident light cone for $d = 25\ \mu$
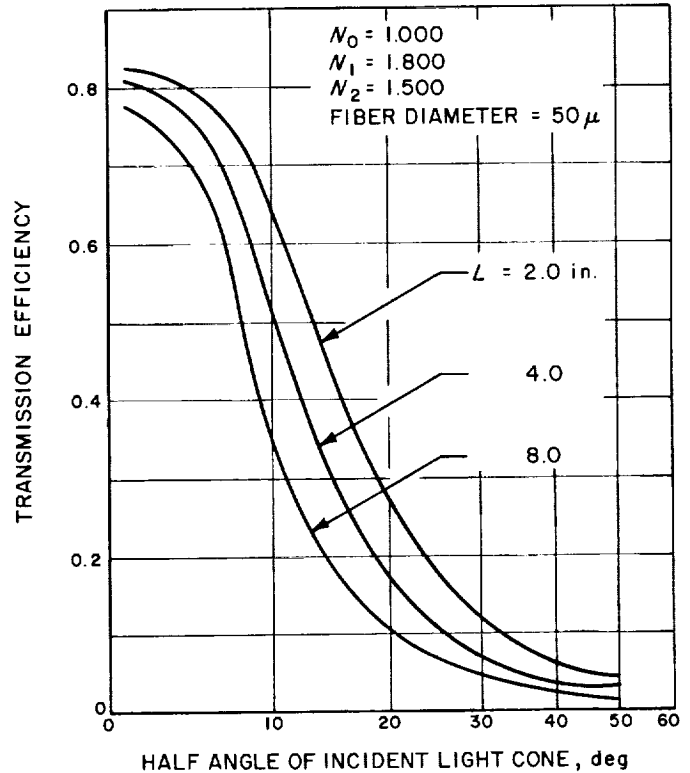
Fig. 10. Transmission efficiency as a function of half-angle of incident light cone for $d = 50\ \mu$

fiber lengths and fiber sizes of 25 and 50 μ, respectively. To illustrate the effects of fiber length on transmission efficiency, curves of transmission efficiency as a function of fiber length for fiber sizes of 10, 25, and 50 μ at a half-cone angle of 8 deg, are shown on Fig. 11. Transmission efficiency, as a function of fiber diameter for 2-, 4-, and 8-in. fibers with a half-cone angle of 8 deg, is shown on Fig. 12.

Figs. 8–10 show that the transmission efficiency drops off very rapidly as the incidence half-cone angle increases. As expected, very little energy is transmitted through the bundle for light rays lying close to and outside the critical angle. The energy transmitted per unit cone angle is small for these rays and the average energy over the entire light cone decreases as the half-cone angle increases. To obtain high transmission efficiency, therefore, the light rays must be transmitted as close to the bundle axis as possible.

Fig. 12 shows the relationship between transmission efficiency and fiber diameter with the fiber length as the parameter. It is seen from these curves that the larger the fiber diameter, the higher the transmission efficiency.

However, the resolving power of the fiber bundle and the resolution of the transmitted image are inverse functions of the individual fiber size. When an image is formed upon the entrance end of the bundle, the multiple internal reflections occurring within the individual fibers tend to integrate any variation in image intensity across the surface of the fiber and the intensity at the exit end of the fiber appears uniformly bright. If the bundle is held stationary with respect to the image, the image is, in effect, sampled by the individual fibers. The resolution of the transmitted image is therefore determined by the fiber size.

3. Conclusions

The analysis of theoretical transmission efficiency of the fiber optic bundle was performed. As a result of the analysis, various numerical data were computed. These data can be utilized in the design of the fiber optic system. In the analysis, it is assumed that the image to be transmitted has already been formed at the entrance surface of the bundle. In a conventional lens system, the surfaces are curved and all rays reaching the lens are
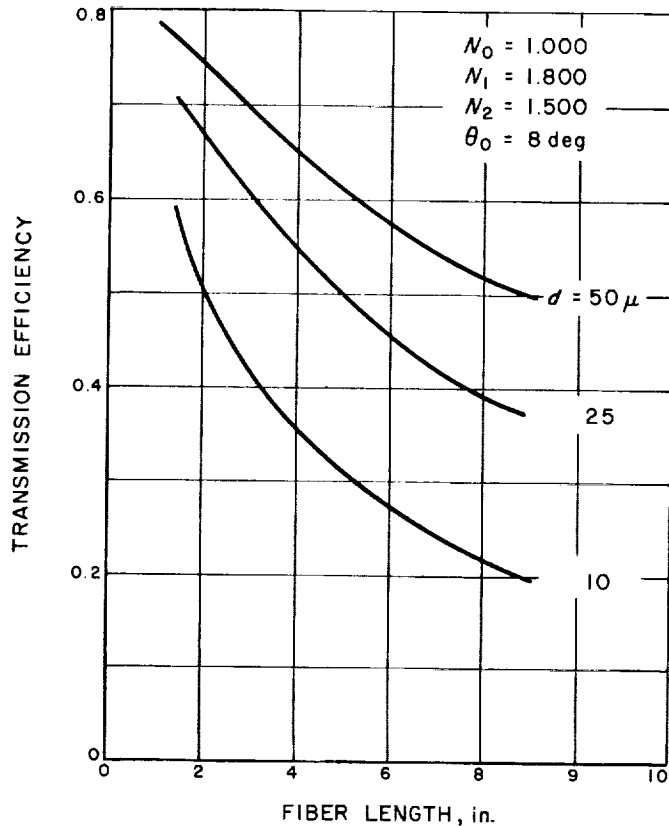
Fig. 11. Transmission efficiency as a function of fiber length
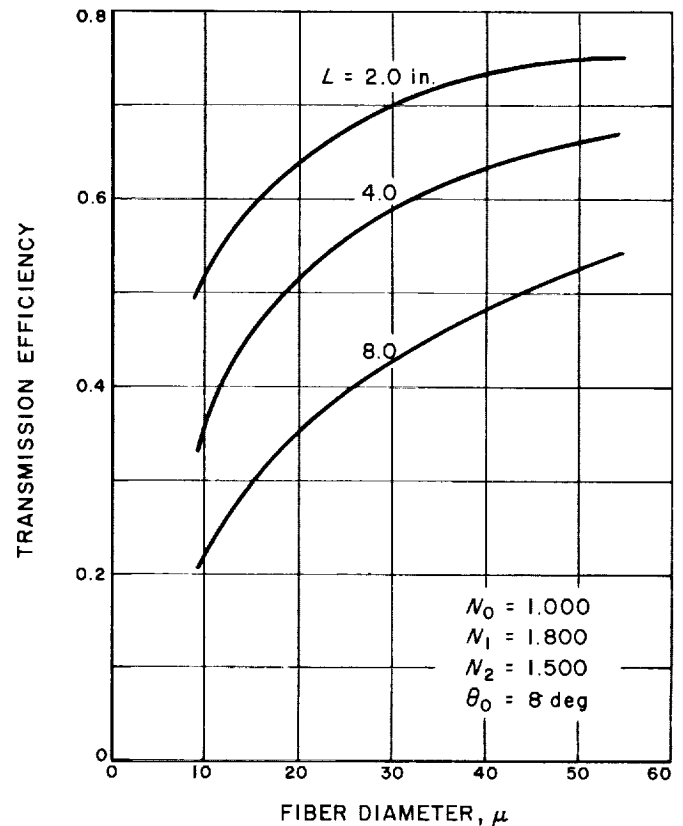
Fig. 12. Transmission efficiency as a function of fiber diameter

bent to meet at a point on the focal plane. The image is formed by the summation of all points on the focal plane. For a fiber optic, no such ordered phase relationship exists. Fiber optics cannot form an image. Therefore, the image must be applied to the entrance surface by an external image-forming device.

A fiber bundle can be designed and fabricated in the form of a cone with the diameter of the entrance end larger than that of the exit end. This fiber cone is capable of transmitting and intensifying an image. There is a net gain in flux per unit area, the numerical value of which

is primarily dependent upon the ratio of the two diameters and the transmission efficiency of the fiber cone. The approximate transmission efficiency of such a cone can be evaluated by Eq. (8) and by assuming that the length of the bundle is equal to the average length of the on-axis and outermost fibers, and that the diameter is the average of the entrance and exit diameters. The fiber cone can be coupled to a simple conventional lens to form a fiber optics system with low f-ratios. The purpose of the cone is to reduce the image size and to allow the utilization of conventional lenses with longer focal lengths and larger effective apertures.

XIII. Applied Science

A. Electron-Excited X-Ray Fluorescence for Lunar Surface Analysis

A. Metzger

1. Introduction

Elements exposed to a suitable source of excitation will emit characteristic X-ray line spectra which can be resolved, according to wavelength, by either electronic or optical discrimination. In the latter case, a crystal–collimator combination is used to diffract a particular wavelength according to Bragg's Law, $n\lambda = 2d \sin \theta$, where λ is the wavelength of the X-ray, d is the interplanar spacing of the diffracting crystal, and θ is the angle between the incident ray and the crystal plane. Only those wavelengths which satisfy the Bragg equation for a given wavelength or higher orders of n will be diffracted. Most K-series X-ray lines are easily detected by gas radiation counters; however, since the mass absorption coefficient of characteristic X-rays increases with decreasing atomic number, detection becomes progressively more difficult for the light elements.

The use of an X-ray spectrograph as a means of performing elemental analysis on powdered samples of lunar material was proposed by Philips Electronic Instruments of Mount Vernon, N. Y. The design approach which emerged from a feasibility study utilized an electron beam, rather than X-rays, for sample excitation in order to minimize the power required for the necessary intensity of emission. Analyzing crystals were chosen as the primary means of wavelength discrimination because of the high resolution of which they are capable. Following the delivery of a breadboard instrument constructed along these lines, extensive studies of X-ray fluorescence induced by electron excitation have been carried out at JPL.

2. Description of Instrument

A schematic of the test instrument is shown in Fig. 1. An electron gun projects a beam on a target surface in
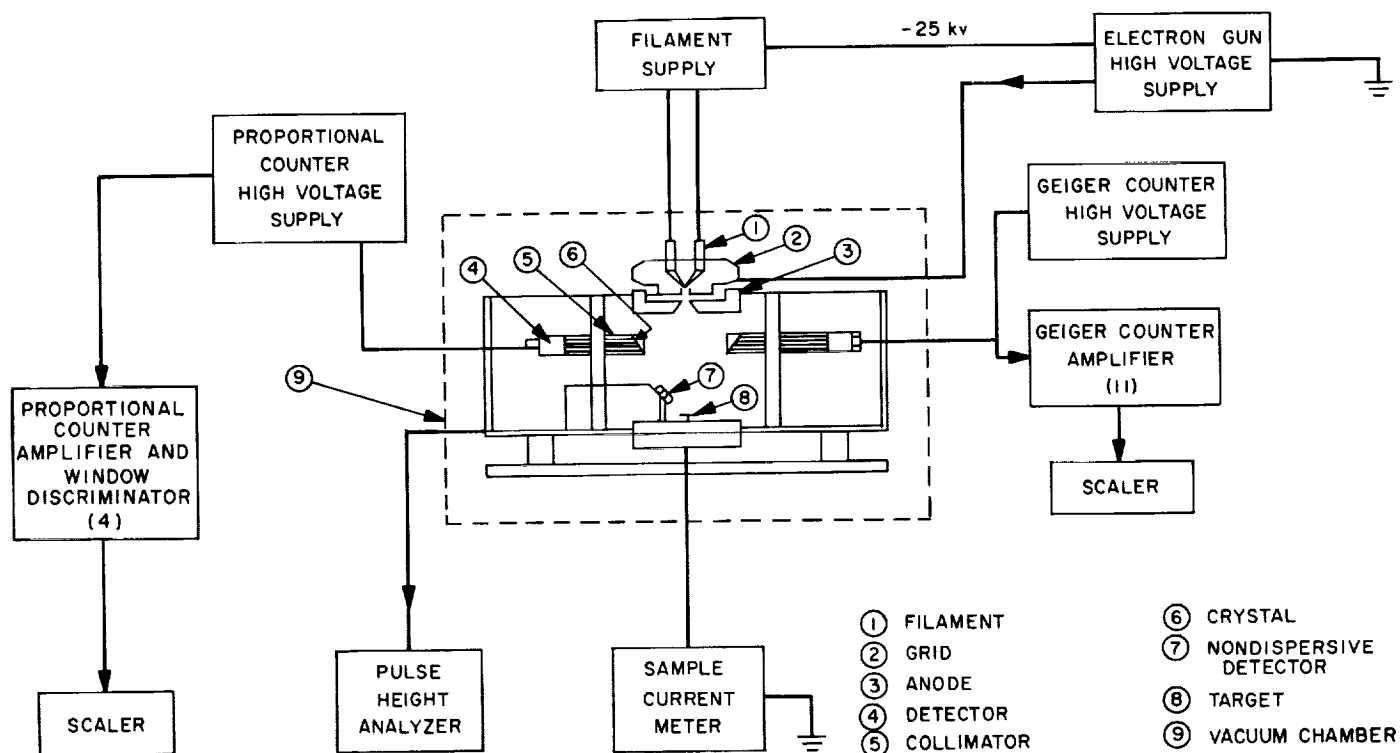
Fig. 1. X-ray spectrograph

the form of a slightly elliptical spot. Fluorescent radiation from the excited sample is viewed by 13 dispersive channels, corresponding to selected elements of geochemical interest. Each channel is composed of a crystal positioned at the reflection angle θ for the appropriate wavelength, a flat-bladed collimator stack which defines the acceptance width $\Delta\theta$ of the diffracted X-rays, and a radiation detector. Geiger counters and proportional counters are used as detectors, the latter for light elements. An additional proportional counter accepts radiation of all wavelengths; its amplified output is transmitted to an analog-to-digital converter for pulse height analysis. Output pulses from the proportional counters in the dispersive channels are passed through a double threshold circuit to improve discrimination against scattered radiation, higher-order diffraction, and possible fluorescence from the crystal.

3. Sample Conductivity

Powdered targets of high electrical resistivity do not behave well under normal conditions of operation. The formation of a negative space charge at the surface of the target reduces the intensity of emission, causes wide fluctuations in counting rate, and, at its worst, violently

disperses the sample. Separate attempts were made to provide a discharge path for the electron beam with a brass mesh placed across the top of the sample, a copper pin at its center, a surface layer of graphite, and graphite distributed homogeneously through the sample. Only the last of these was entirely satisfactory. Measurements showed that the addition of 5 to 15 wt % of graphite lowered the electrical resistivity of nonconducting materials by 5 to 7 orders-of-magnitude. Precharging the sample preparation chamber with a small quantity of conductive material will stabilize the response of the lunar specimen to the electron beam.

4. Target Temperature

Bombardment by 10^{15} electrons/sec with an average energy of 25 kev delivers an energy of 4×10^7 ergs/sec to the target, much of which must be dissipated as heat. In order to determine if standard operating conditions produce an unacceptable level of sample heating, the temperature of a powdered target has been measured as a function of beam flux and spot size by means of a thermocouple embedded in a 200-mesh sample of granite. Temperatures measured ranged from 44 to 526°C. For the electron gun conditions considered as standard, 25 kv

of potential, 50 μa of current, and a spot diameter of $3/16$ in., the temperature was 234°C, eliminating any cause for concern.

5. Specimen Presentation

a. Procedure. Material for powder tests is prepared by grinding, if necessary, and sifting to eliminate particles above a given size. The sample is prepared by weighing and is mixed by tumbling and shaking in a plastic vial. A graphite cup with a cavity of $1\frac{1}{8}$-in. D and $1/16$-in. depth is filled with an excess of powder. The powder is leveled off and compressed with a flat metallic surface. Excess powder is simply decanted, and the specimen is ready for use.

The possibility of preferential sorting during grinding, shaking, or tumbling has been investigated for five sample mixtures representing a variety of particle characteristics. A combination of tumbling and shaking was found adequate to prevent segregation.

b. Compaction and surface smoothness. Fluorescent emission has been studied as a function of sample compaction. The intensity from powdered iron increases less than 5% between 5 and 1000 psi. Aluminum oxide presents a sharp contrast, showing more than a 60% increase in intensity over the same interval, with almost three-quarters of the change occurring below 100 psi. The principal effect of compaction, particularly at low pressures, is to improve the surface smoothness of the target. This reduces absorption losses by decreasing the effective emergent path length of the X-rays. Since the characteristic X-rays of aluminum have a longer wavelength and are therefore more easily absorbed than those of iron, the aluminum intensity should be more sensitive to changes in surface smoothness.

A second factor involved in determining the effective path length of fluorescent radiation is the takeoff angle, i.e., the angle which the emitted radiation makes with the surface of the sample while undergoing dispersive diffraction. When the takeoff angle for aluminum radiation was increased $2\frac{1}{2}$ times, the effect of compaction was reduced by a factor of 4, thereby removing any excessive dependence of intensity on surface condition.

c. Consolidated targets. Although the fixed-channel lunar X-ray spectrograph was designed for the analysis of powdered samples, preliminary tests have been made

of the response characteristics of nonconducting consolidated targets, with a painted Aquadag trail or metal finger providing the discharge path for the electron beam. Stable fluorescent emission has been obtained from rock samples and from a piece of glass. This interesting line of investigation is continuing to determine how much analytical capability can be retained if sample processing is eliminated.

6. Effects of Heterogeneity

The effect of particle size and shape on the intensity of response has been investigated with homogeneous and heterogeneous samples. The same class of matrix effect reported for fluorescent excitation (Ref. 1) has also been found to exist with electron excitation. For a sufficiently small mean particle size of the element to be detected, fluorescent emission is enhanced at low concentrations, so that the intensity is no longer linearly proportional to concentration. Although particle sizes of a micron or less effectively increase the sensitivity at low concentrations, it is better to avoid excessive grinding and retain a linear response. Linear results with unground iron powder indicate that this can be obtained without going above $75-\mu$ (200-mesh) particles, which still allows a sufficiently smooth target surface to be retained. This heterogeneity effect suggests that the calibration standards, whether chemical compounds or pre-analyzed rocks, should have particle characteristics similar to those of rock material.

7. Analytic Response

a. Dispersive analysis. The interelemental effects encountered in electron-excited fluorescent emission have been studied. These include the absorption of emitted X-rays in the specimen, generation of an X-ray continuum due to deceleration of the incident electrons in the target, and secondary emission initiated by the primary fluorescent response. Such effects have been observed in certain multicomponent calibration curves. Sets of calibration curves for calcium and aluminum are given in Figs. 2 and 3, respectively. The calcium curves are virtually identical. On the other hand, the characteristic X-rays of aluminum are strongly absorbed by iron and enhanced by calcium. The additive nature of the aluminum calibration curves is significant for data interpretation.

Responses obtained from these multicomponent mixtures have been used as calibration curves in the analysis of unknown mixtures containing the same compounds. The
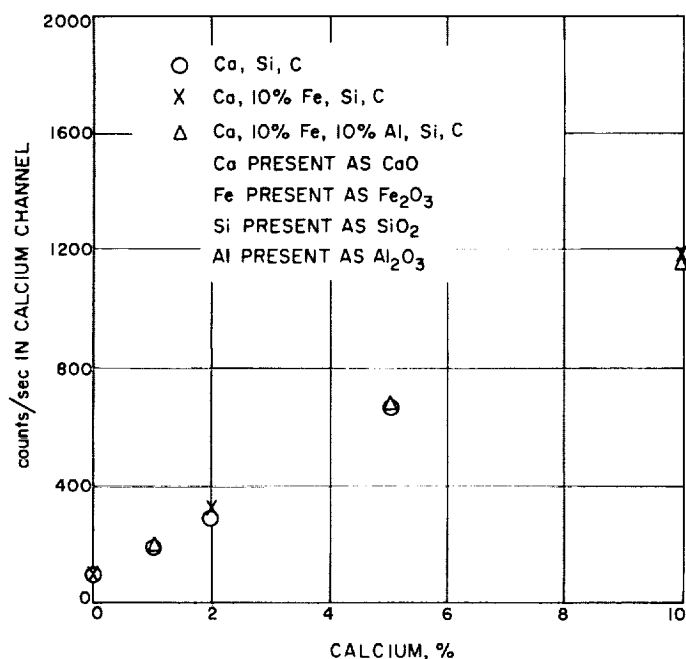
Fig. 2. Calcium calibration curves—dispersive

Fig. 3. Aluminum calibration curves—dispersive

results for two unknowns are shown in Table 1. They indicate that the instrument is capable of performing accurate quantitative measurements on multicomponent samples.

b. Nondispersive analysis. The capability of nondispersive analysis to provide quantitative information has been tested by obtaining the spectral distribution of a set of elements as standards and then rerunning the same specimens additively to produce composites. The composite spectra were stripped using a computer program developed by Dr. Jacob Trombka (Ref. 2). The results are shown in Table 2. Although these composite tests were entirely independent of sample effects, they indicate that sensitivity levels of 1 to 5% appear reasonable, even for complex cases. The promising gross analytical capability

of nondispersive analysis deserves consideration as a separate instrument, less sensitive but simpler than a dispersive instrument, particularly if direct surface excitation is found to be practical.

c. Rocks. Preliminary tests of the quantitative capability of X-ray fluorescence electron excitation have been made with analyzed rock specimens. The results from four dispersive channels are plotted in Fig. 4 as a function of their predetermined composition. No background corrections have been applied to the observed counting rates. The response of iron in granite suggests that a spectroscopic analysis of the specimen tested will show a higher iron concentration than presently indicated. The intensities of Fig. 4 are clearly proportional to composition. The results for titanium demonstrate that the sensitivity of the spectrograph for the analysis of rocks extends well below 1%.

Table 1. Results[a] of dispersive analysis of quaternary unknowns: percentage of element in sample

Element	Sample 1			Sample 2		
	Measured	Pre-pared	Difference	Measured	Pre-pared	Difference
Fe	7.5	7.9	−5	3.4	3.6	−5
Ca	3.5	3.6	−3	7.9	8.1	−2
Si	40.0	37.4	+7	35.8	36.4	−2
Al	1.7	1.9	−10	3.9	3.0	+30

[a]The Ca results were revised after reassessing the calibration curves; the Si results were calculated by difference.

Table 2. Results of nondispersive composite tests[a] with computer analysis[b, c]

Run	Mg	Al	Si	S	Ca	Ti	Cr	Mn	Fe	Ni
YI	0.74	1.08	——	——	0.015	0.0023	——	——	——	0.00097
ZI	——	——	1.23	0.83	——	0.00006	——	——	——	——
AJ	——	——	——	——	0.88	1.08	——	——	——	——
BJ	——	——	——	——	——	0.0011	——	——	0.98	1.00
CJ	0.96	——	——	——	——	0.013	——	——	——	0.97
DJ	0.805	——	1.00	0.971	——	0.0064	0.0078	——	——	——
EJ	——	——	——	——	——	0.939	1.16	0.734	——	——
FJ	——	——	0.0055	——	——	——	1.03	——	0.89	1.03
GJ	1.05	——	1.29	0.88	1.08	0.93	——	0.17	0.77	——
HJ	0.57	1.31	0.79	——	1.03	——	——	0.099	0.91	——

[a] With a 25-kv, 4-μa electron beam; side window proportional counter, P10 fill.
[b] ░░░ numbers represent those elements which made up the composite.
[c] If the element is present, the correct relative intensity is 1.00; if the element is absent, 0.0.
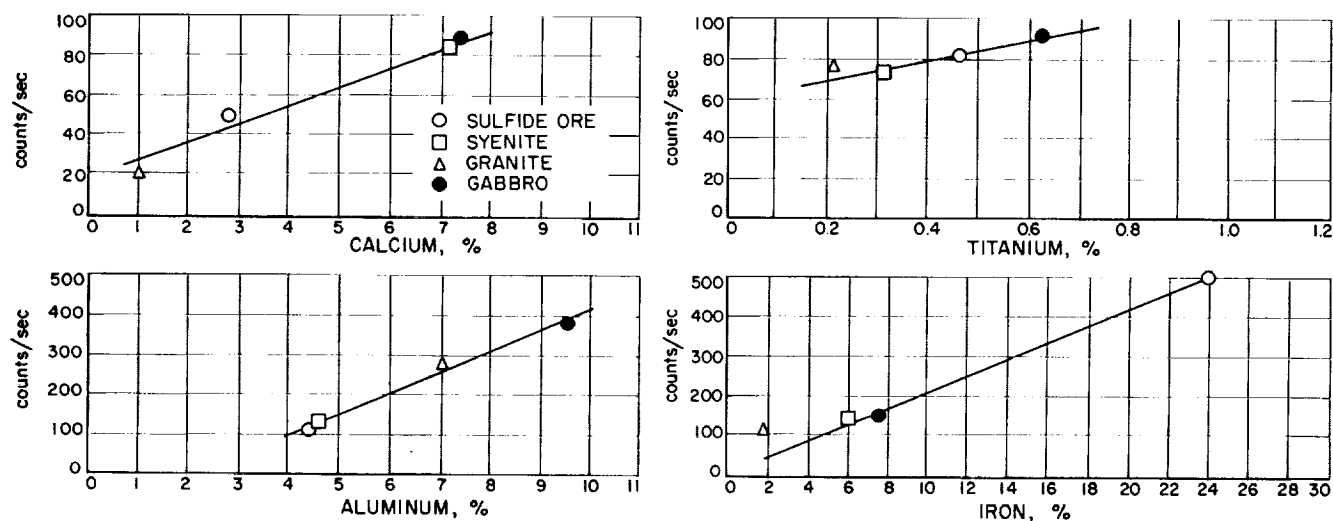
Fig. 4. Measured intensity vs composition of rock specimens

8. Summary

The significant results of this phase of experimentation with the X-ray spectrograph are as follows:

(1) Stable conditions of excitation and detection have been obtained.

(2) No need has been found for resorting to unreasonable requirements such as extreme compaction pressures or sample fusion.

(3) Heterogeneity effects can be significant and are closely related to methods of sample preparation.

(4) A nondispersive pulse height analysis system complements the dispersive capability of the instrument.

(5) Reproducible and internally self-consistent results have been obtained in systematic measurements with three and four component specimens.

(6) Initial experiments with rock specimens show that the instrument is capable of quantitative analysis for elements present in major and minor abundance.

B. Analysis of the Martian Atmosphere by α-Particle Bombardment— The Rutherford Experiment

J. H. Marshall and E. J. Franzgrote

The recent work of Kaplan, et al., (Ref. 3) indicates that about 40 to 90% of the Martian atmosphere cannot be identified spectroscopically. N_2 and Ar, which are suspected to be two of the major constituents, cannot be measured by present Earth-bound techniques. The lack of exact knowledge about the composition and density of the Martian atmosphere produces questions relevant to Martian biology and geology, and also adds to the difficulties in landing a sophisticated capsule on Mars.

One solution to the above problem is to drop a relatively simple capsule through the Martian atmosphere and measure its composition directly. A technique applicable to such a capsule is described here. This experiment, which is designed to measure the three suspected major constituents (Ar, N_2, and CO_2), is characterized by low weight, power, and data rate.

1. Analyses

Ernest Rutherford's studies of the scattering of α-particles by various materials led to his theory, in 1911, that the charge of atoms is concentrated in small central nuclei. In 1919 Rutherford discovered that long-range particles were produced by α-particle bombardment of air, and he correctly interpreted his findings as being due to the disintegration of N_2 nuclei. These two phenomena, which led to the discovery of the atomic nucleus and to its artificial transmutation, form the basis for the analytical technique described here.

a. N_2 analysis. The disintegration of N_2 by α-particle bombardment with the resultant production of protons, $N^{14}(\alpha, p)O^{17}$, is sometimes called the Rutherford reaction. A number of lighter elements undergo the (α, p) nuclear reaction with α-particles of 6 Mev. However, N^{14} is the only major isotope of this type present in planetary atmospheres; C^{12}, O^{16}, and Ar^{40}, for example, are stable to 6-Mev α-particles. Fig. 5 shows proton spectra obtained from several gases at a pressure of 70 mm Hg. Simple counting of protons between 2 and 4 Mev gives a specific measurement for N_2. CO_2 and Ar (Fig. 5a and b) contribute less than 1% as many counts as N_2 (Fig. 5c) contributes in the region of the N_2-proton peak. A sample
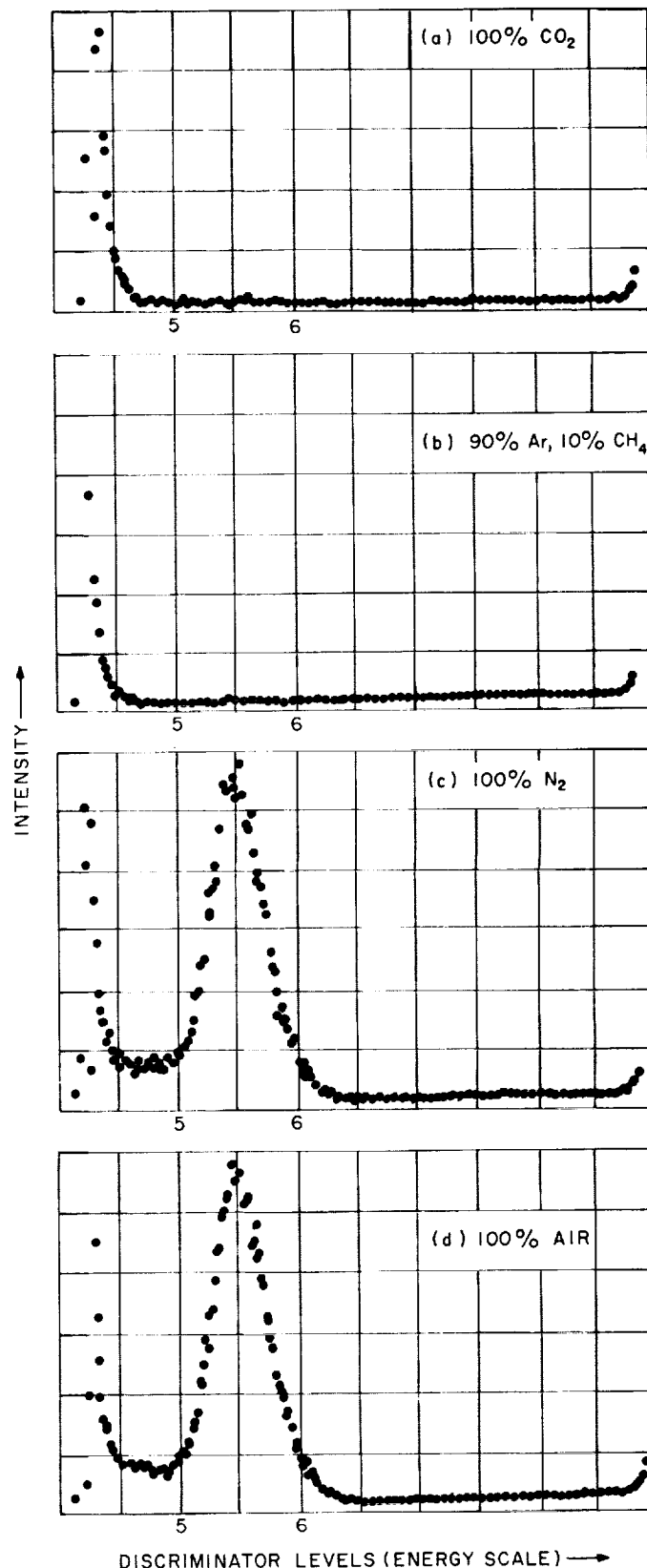
Fig. 5. Proton energy spectra

of air has been analyzed using this method (Fig. 5d). The percentage of N_2 in air was found to be $80 \pm 2\%$.

b. CO_2 and Ar analysis. Coulomb scattering of the positively charged α-particles by nuclei is generally called Rutherford scattering. The energy of a scattered α-particle is a function of its initial energy, the angle of scattering, and the mass of the target nucleus. Thus, the energy spectrum of scattered particles characterizes the mass of target nuclei if the initial energy and the scattering angle are fixed. For 180-deg backscattering,

$$E_s = E_0 \left(\frac{A - 4}{A + 4} \right)^2 ,$$

where E_s is the energy of the scattered particle; E_0, the initial energy of the particle; and A, the mass number of the target nucleus (mass of the α-particle $= 4$).

Fig. 6 shows energy spectra for backscattering from CO_2, N_2, and Ar at a pressure of 70 mm Hg. Four discriminator levels could be set, as indicated, to measure these gases. Because of the relatively large mass of the Ar^{40} nucleus, Ar can be measured by counting pulses between Levels 3 and 4 without interference from the other gases. N_2 would contribute some counts in the region of the C peak, but this contribution can readily be calculated from the specific (α, p) measurement of N_2.

A redundant measurement for N_2 and CO_2 may be made by counting pulses due to N^{14} and O^{16} in the intermediate range between Levels 2 and 3.

2. Description of Instrument

a. Source–sample–detector design. The instrument would be similar in some respects to the *Surveyor* α-particle scattering instrument proposed by Professor Anthony Turkevich of the University of Chicago for measurement of the composition of the lunar surface. The sample is exposed to 6.1-Mev α-particles from Cm^{242}, and solid-state detectors are used to count scattered and produced particles.

A suggested mechanical layout of the instrument may be seen in Fig. 7. The sample volume may be simply an enclosure into or through which the atmosphere is allowed to flow. A high-angle geometry will be used for detecting scattered α-particles. The proton detector will be located at an angle of about 45 deg to the source. Scattering from walls of the enclosed sample volume can be kept at a minimum by proper choice of materials and geometry of construction.
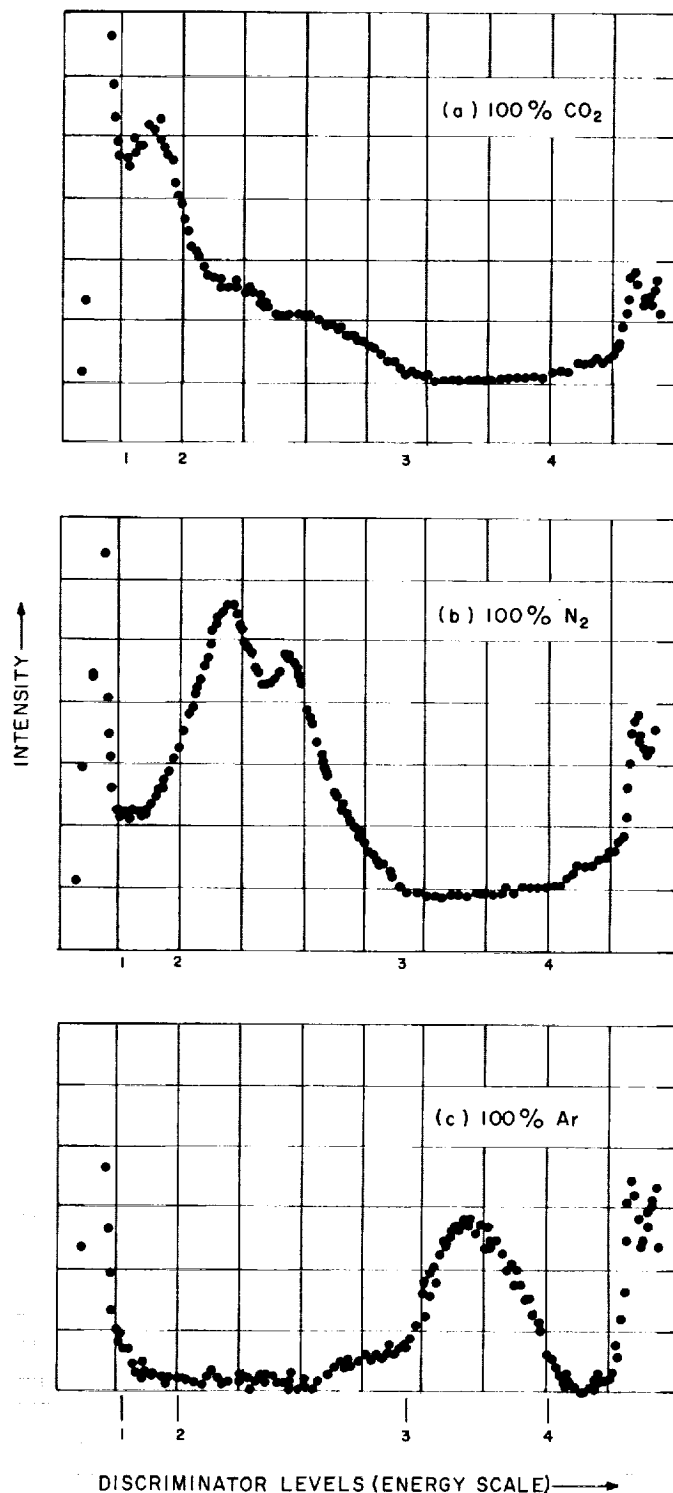
Fig. 6. α-Particle scattering energy spectra

Two solid-state detectors of about 1-cm² area would be used. The proton detector would be covered with a thin foil to stop the 6.1-Mev α-particles but pass the longer
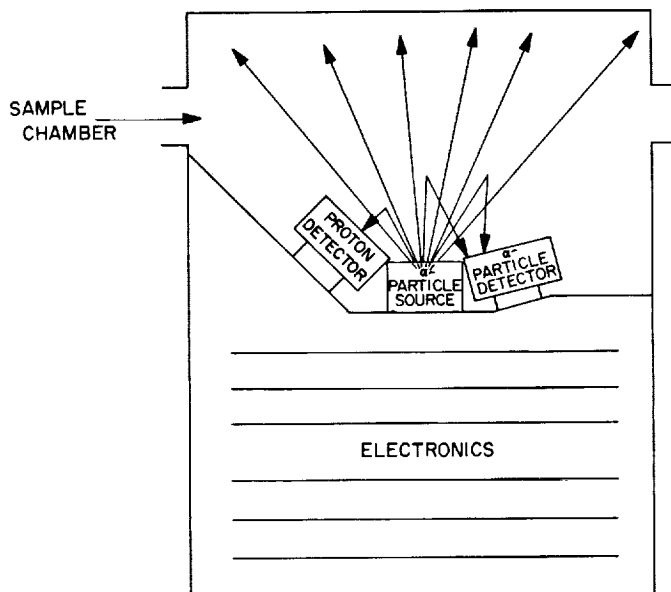
Fig. 7. Mechanical layout of source, sample, and detector

range protons. Some solid-state detectors have survived heat sterilization, but it is not known at the present time how soon this may be done reliably. This problem is now being investigated.

The two main factors that determine the choice of the required α-particle source are its energy resolution and its intensity at Mars. It can be shown that the optimum combination of these two factors is obtained with any isotope whose half-life is shorter than the life of the mission. Cm^{242} with a half-life of 163 days appears to be a logical choice. Dr. James H. Patterson of Argonne National Laboratory has prepared a Cm^{244} source for use in a breadboard model of this instrument. Such a source, which is equivalent to 1 C of Cm^{242}, produces about 1000 events/min from a component with a partial pressure of 20 mbar.

b. Electronics. A block diagram of the electronic circuits proposed for this experiment is shown in Fig. 8. The detector output pulses are first amplified and shaped by charge-sensitive amplifiers. Both detectors have separate amplifiers, whose outputs are connected to several discriminators. The amplifier gains are estimated from presently available breadboard models to be stable to ±0.3% for temperatures between 50 and −50°C.

The discriminators, which are univibrators (also called single-shots or monostable multivibrators), trigger when their input pulse exceeds a preset bias level. Because each bias voltage corresponds to a given α-particle or proton energy, the discriminators determine whether a

given particle did or did not exceed a preset energy. The four discriminators connected to the output of the α-particle amplifier define the four boundaries of the three α-particle energy channels. Similarly, the two discriminators connected to the proton amplifier define the two boundaries of the proton energy channel. Typical discriminator bias levels and output pulse widths are shown in Table 3. The discriminator thresholds for presently existing breadboard models are stable to ±0.2% over the temperature range from −50 to 50°C.

Table 3. Typical discriminator bias levels and output pulse widths

Discrim-inator	Pulse width, μsec	Bias, Mev	Function: definition of boundary
1	10	0.6	Lower boundary of CO_2 channel
2	12	1.18	Upper boundary of CO_2 channel and lower boundary of redundant channel
3	10	3.0	Lower boundary of Ar channel and upper boundary of redundant channel
4	14	5.0	Upper boundary of Ar channel
5	10	2.0	Lower boundary of N_2 channel
6	12	4.0	Upper boundary of N_2 channel

Four coincidence circuits sort the discriminator output pulses for a given event into one of four channels, corresponding to the four energy windows, or determine that the event is not to be counted. Each coincidence circuit consists of an *and* gate with a capacitor to slow the output fall time and a buffer to produce an output pulse when the input signal falls half way to its final value. Typical pulses for the C channel are shown in Fig. 9, which illustrates the case for the energy marginally large enough to trigger Discriminator 2. Because the discriminators trigger within about 0.3 μsec after the input pulse exceeds the bias, Discriminator 1 may trigger as much as 2 μsec before Discriminator 2. In order to prevent the biased buffer from giving an output pulse in this case, a capacitor slows the *and* gate fall-time so that 4 μsec must pass before the biased buffer is triggered. For the same reason the anticoincidence pulses are made longer than the coincidence pulses. From Fig. 9 it can be seen that, if Discriminator 2 triggers at the latest possible time, then the *and* gate output voltage falls only half way to the trigger level of the biased buffer. On the other hand, if Discriminator 2 does not trigger at all, then the biased buffer gives a 6-μsec output pulse, which is counted by the remainder of the digital circuits. The *and* gates may be closed by the count gate,
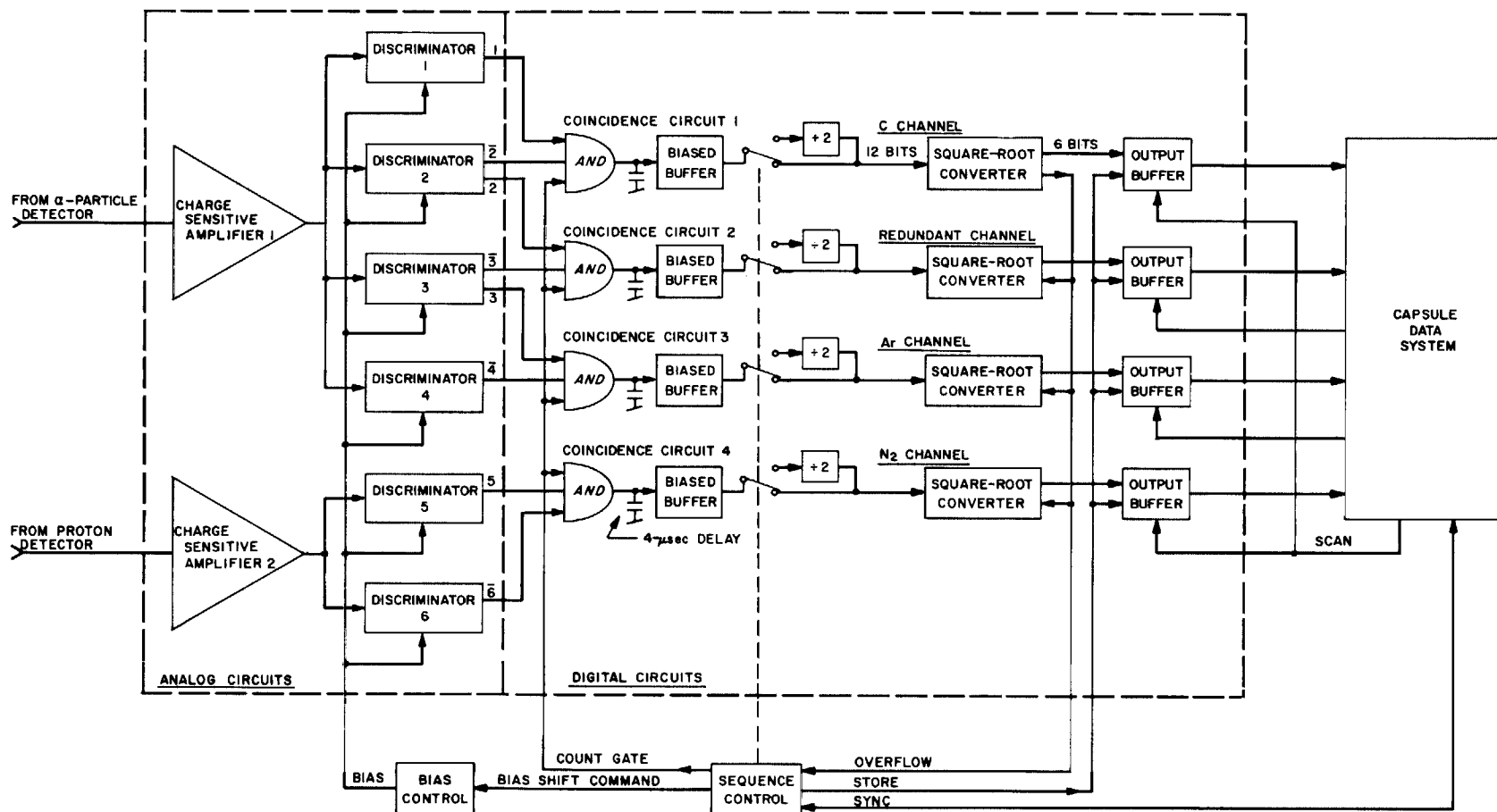
Fig. 8. Electronic circuits proposed for the Rutherford experiment
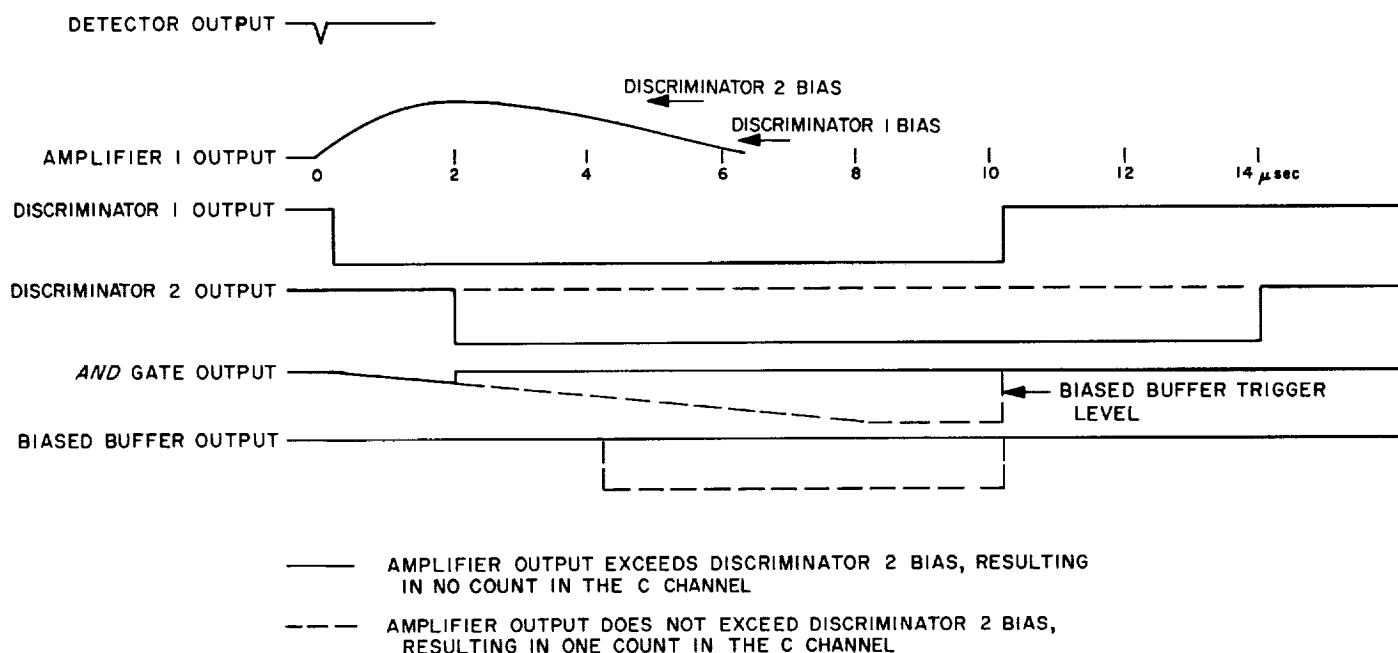
Fig. 9. Timing diagram for the acceptance or rejection of a count by the C channel

which inhibits further counting during the readout and reset of the buffers.

The coincidence circuit output passes through a flip-flop which, if so commanded by the sequence control, will divide the counting rate by two. Usually this flip-flop is bypassed, and the coincidence circuit output enters the square-root converter directly. However, if the square-root converter has overflowed on the previous run, then this flip-flop divides the counting rate by two to prevent overflow on the next run.

The square-root converter[1] output equals approximately the square root of the number of input pulses N within an rms rounding error of approximately $\pm(N/3)^{1/2}$. In this way, the number of bits for this experiment is reduced by a factor of two, while the total rms error is increased by only 16% above the statistical error. Each converter consists of two 6-bit registers with comparison gating and a single binary stage driven by the input pulses. One register counts the single binary output until the binary number in the first register equals that in the second register. Then, one count is added to the second register, and the first register is reset. The square of the number stored in the second register equals $(N - \frac{1}{2})$, with a maximum quantization error of $\pm(N - \frac{1}{2})^{1/2} + \frac{1}{2}$.

---

[1] A detailed description of the square-root converter design will appear in the next issue of *SPS*, Vol. IV.

On command from the sequence control, the contents of the second register of the square-root converter are stored in the 6-bit output buffer. When the capsule data system is able to transmit the data back to Earth, it scans this buffer.

The sequence control determines the time during which the various functions of this experiment are carried out. In the case of the most tenuous atmospheres considered, about 145 sec would elapse from the opening of the capsule parachute until transmitter "lock" is achieved. During the next 50 sec, data might then be transmitted to Earth at the rate of 3 bits/sec, giving a total of 150 bits. For such a capsule, the sequence of operation is as follows:

Before the capsule enters the Martian atmosphere, the count gate is opened and calibration data are taken. The count gate is then closed, and these data are transmitted back to Earth. All the flip-flops are reset, and the count gate is left closed in preparation for an atmospheric measurement.

Just after the parachute is deployed, the count gate is opened and counts are accumulated for 2 min. At the end of this time, the count gate is closed, and the outputs of the square-root converters are stored in the output buffers. After all flip-flops (except those in the output buffers) are reset, the count gate is again opened and

counts are accumulated for an additional 1 min. During this time transmitter lock should be achieved, even in the worst case, and the data from the first run will be transmitted back to Earth. If impact has not occurred by the end of the 3 min after parachute deployment, then the data from the second run are stored in the output buffers and transmitted back to Earth. This sequence, in which runs lasting 1 min are stored and then transmitted, continues until impact. For the worst-case descent profile, sufficient time should be available for the first run, but not for any additional runs. For a nominal 20-mbar atmosphere, several runs would be possible. This experiment will use 24 of the 150 available bits in the first 50 sec; for a longer descent time, additional data can be used at the rate of 0.4 bit/sec until impact. This sequence could easily be varied to adapt to other descent profiles.

The sequence control could be designed to automatically command, via the bias control, a change in the discriminator bias levels after the first or second run. This would be done to obtain additional spectral information as a check on proper operation and on the possibility that additional gases are present.

If any square-root converter overflows by filling its second register, the count gate will immediately be closed, preventing the accumulation of further counts in any of the channels. Such an event will be identified by setting all output bits for the channel with overflow to *one*. For this run the measurements of relative concentrations will still be valid, but the absolute concentrations will be unknown because the accumulation time will be uncertain. For subsequent runs, the scale-of-two in front of the square-root converter which overflowed will be activated. Overflows should occur only if the surface pressure exceeds 80 mbar.

c. *Weight and power allocations.* Preliminary estimates of the weight and power allocations for this experiment are given in Table 4. The digital circuits are included in

Table 4. Preliminary weight and power allocations

Unit	Power, mw	Weight, lb
Analog circuits	240	0.5
Digital circuits and sequence control	235	1.0
Power supply	360	0.5
Mounting for source and detectors	—	0.5
Total	835	2.5

the estimates, although a part of these circuits may eventually be shared by several experiments. The power supply for the analog circuits is assumed 50% efficient, because these circuits require voltage regulation of about $\pm 1\%$ for stable operation. Because the digital circuits require only $\pm 20\%$ voltage regulation, the supply for them is assumed 75% efficient.

3. Accuracy

The accuracy of this experiment is limited by several factors, the effects of which will be estimated here. The shapes of the α-particle and proton energy spectra are a weak function of gas pressure because of energy losses resulting from ionization. This energy loss appears not only as a reduction in detected particle energy, but also as a spectral shift caused by exciting nuclear resonances at lower energies. Fortunately, at the low pressures and path lengths expected for this measurement, α-particles traversing the sample volume will lose very little energy except through nuclear collisions. Effects from ionization energy loss will be small and subject to accurate calibration, depending only on geometry and partial pressures. Data for this correction can be obtained on Earth by measuring simulated atmospheres. The count rate will be a single-valued function of, and nearly proportional to, the number density of the measured nuclei. Thus, not only the relative concentrations of the three gases may be determined, but also their absolute densities. The accuracy, as with any density measurement, depends upon how well the pressure and the temperature of the sample volume are known relative to ambient conditions. This measurement of density as a function of altitude could be made independent of other measurements by assuming an average scale height and extrapolating back from impact. More accurate results could be obtained if altimeter readings were also available. Readings of the count rate as it increases during descent of the capsule would serve as a valuable check on both the performance of this instrument and the validity of concurrent pressure and temperature measurements.

If N_2, Ar, and CO_2 were not the major constituents of the atmosphere, the data would be subject to erroneous interpretation. However, sound theoretical and experimental evidence can be given for their presence, and spectroscopic upper limits of less than 1% have been set for other gases such as O_2, NO_2, CO_2, CH_4, and NH_3 (Refs. 3–6). The redundant α-particle scattering measurement of nuclei of mass 12 to 40 would provide some check on this assumption.

The effects of cosmic ray and solar proton background must also be considered. Because only charged particles in limited energy ranges would be detected, cosmic rays would produce a negligible background of about 1 count/min for the detector areas and energy windows under consideration. Typical solar flares, on the other hand, have intensities many orders-of-magnitude above the cosmic ray level. Solar flare protons below 100 Mev would be stopped by the overlying atmosphere and capsule material. Using solar flare data for the period from 1956 to 1961, and choosing a tolerable limit of about 10 counts/min, the probability of avoiding flares is 94 to 96%. This figure could be somewhat improved by the use of some shielding at the detector. Since typical $1/e$ rise and decay times for flares are several hours, a background calibration shortly before planetary entry would be valid.

Drifts in the discriminator triggering levels and amplifier gains cause changes in the widths of the energy windows and thus in the efficiency for detecting α-particles or protons from a given element. Through judicious choice of the boundaries of the energy windows, these drifts should cause errors of less than $\pm 1\%$ in the concentrations. The calibration before entering the atmosphere would verify whether or not this stability was actually achieved.

The statistical error, which is proportional to the square root of the number of counts received, can be calculated from Poisson statistics and from the known rounding error of the square-root converter. The total rms error in percent is then given by

$$\sigma = \left(\frac{13,300}{N} + 1 \right)^{\frac{1}{2}}$$

where N is the total number of counts received for either N_2 or Ar. In the case of CO_2, an additional error arises since the CO_2 counts must be corrected for background counts produced by N_2. If the counting rate for a component with a partial pressure at the surface of 20 mbar (about 3×10^{-5} g-cm^{-3}) were 1024 counts/min, then N and σ for N_2 or Ar during the first run would be as follows for surface pressures P_s of 10, 20, and 80 mbar:

P_s, mbar	N	σ, %
10	512	5.2
20	1024	3.8
80	4096	2.1

The concentration of CO_2 is most susceptible to Earthbound measurements (Ref. 3) and may possibly be measured quite accurately in several years. A comparison of Earth-bound CO_2 measurements with the results of this experiment would provide a check on the proper functioning of the equipment.

A scientific breadboard model for this experiment is nearly completed and should provide more accurate estimates of errors and general feasibility.

References

1. Claisse, F., and Samson, C., "Heterogeneity Effects in X-Ray Analysis," *Advances in X-Ray Analysis,* Vol. 5, pp. 335–354, Plenum Press, New York, N.Y., 1962.

2. Trombka, J. I., *Least-Square Analysis of Gamma Ray Pulse Height Spectra,* Technical Report No. 32-373, Jet Propulsion Laboratory, Pasadena, California, December 15, 1962.

3. Kaplan, L. D., Munch, G., Spinrad, H., "An Analysis of the Spectrum of Mars," *Astrophysical Journal,* Vol. 139, No. 1, January 1, 1964.

4. Spinrad, H., "The NO_2 Content of the Martian Atmosphere," *Publications of the Astronomical Society of the Pacific,* Vol. 75, No. 443, p. 90, April 1963.

5. Dollfus, A., and Lyot, B., *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences,* Vol. 229, p. 1277, 1949; Vol. 231, p. 1430, 1950; and Vol. 232, p. 1066, 1951.

6. de Vaucoulers, G., *l'Astronomie,* Vol. 65, p. 103, 1951; and *Astronomical Society of the Pacific Leaflet No. 276,* 1952.

XIV. Chemistry

A. Use of a Varian V-4210A RF Unit as a Heteronuclear Spin Decoupler

D. D. Elleman, C. D. Pearce, and S. L. Manatt

The Varian V-4210A RF unit is a variable-frequency RF transmitter–receiver designed for wide-line nuclear magnetic resonance studies from 2 to 16 Mc. Its output is about 4 to 6 w. The stability of this RF source is only about ± 20 to 100 cps/min, depending on the frequency range selected. We would like to describe a simple addition to the V-4210A unit which not only stabilizes the frequency, but permits frequency modulation of the RF source as well.

Fig. 1 shows a crystal-controlled oscillator linked to a simple circuit for frequency modulation. The parameters for such an oscillator depend on the frequency desired; the parameters for the audio frequency part are given. We have found it convenient to add a variable capacitor
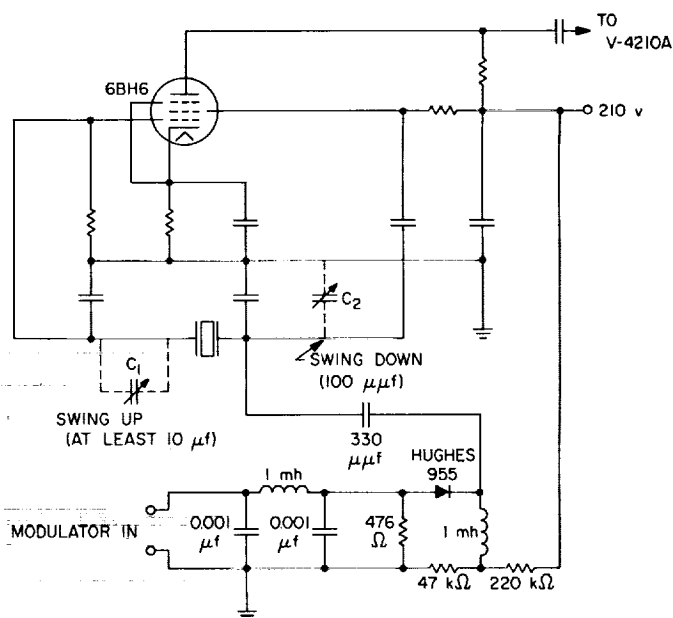
Fig. 1. Crystal-controlled oscillator with provision for frequency modulation with a V-4310A RF unit

C_1 to swing the center-band frequency up and a variable capacitor C_2 to swing it down. In this manner, it was feasible to swing the frequency up about 1000 cps and down about 200 cps; this adjustment could be made to better than 0.1 cps.

Several crystals which differ in frequency by about 1000 to 2000 cps can be used to achieve a wider range. The output from the crystal-controlled oscillator is fed to the crystal jack on the front of the V-4310A RF unit. Frequency stabilities of better than 0.1 cps/hr were obtained. One can thus do high-resolution spectroscopy with this unit if desired.
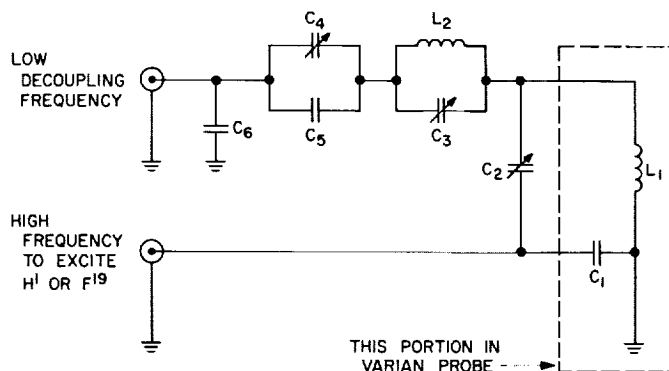
Fig. 2. Circuit for double tuning a Varian nuclear magnetic resonance probe
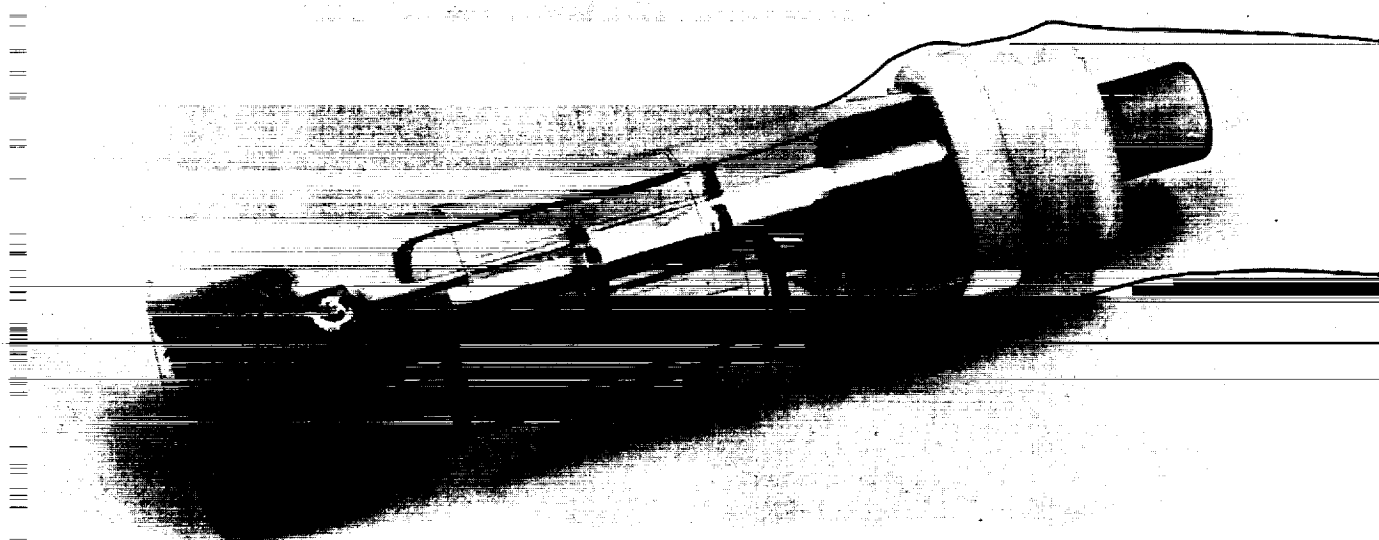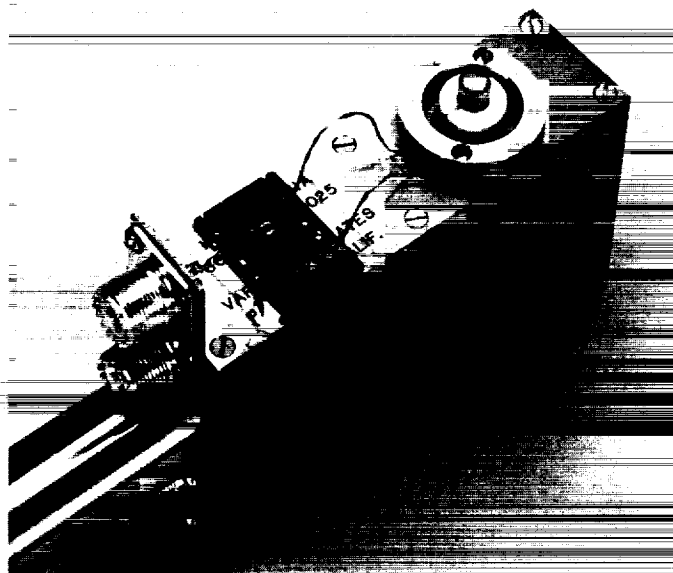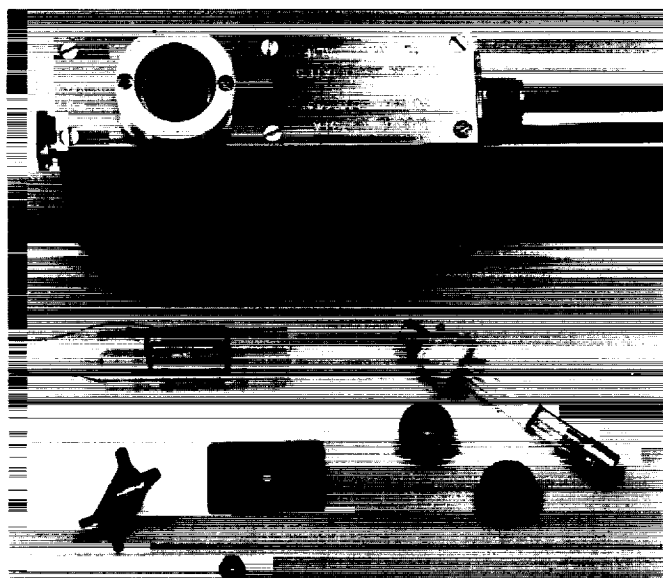
Fig. 3. Construction and assembly of a second transmitter coil for a Varian nuclear magnetic resonance probe

For homonuclear decoupling, the strong irradiating fields can be introduced as audio sidebands. For heteronuclear decoupling, the usual method for introducing two RF frequencies at the sample is to double tune the probe transmitter coil. Since the means for doing this is not well-known, such a scheme is shown in Fig. 2. A second method which works satisfactorily and has several advantages is to construct a second transmitter coil on a glass form, which is then placed coaxially with the probe receiver insert. The construction and assembly of this

coil are shown in Fig. 3. The details of the construction for this modification will be described elsewhere. Figs. 4 and 5 show two examples of the performance of such a coil system when the second RF is derived from a V-4310A unit with the crystal oscillator described above.

These results indicate the resolution is not noticeably reduced by the presence of an additional coil. The proton resonance spectra were obtained with the field-frequency-lock spectrometer described previously (Ref. 1).
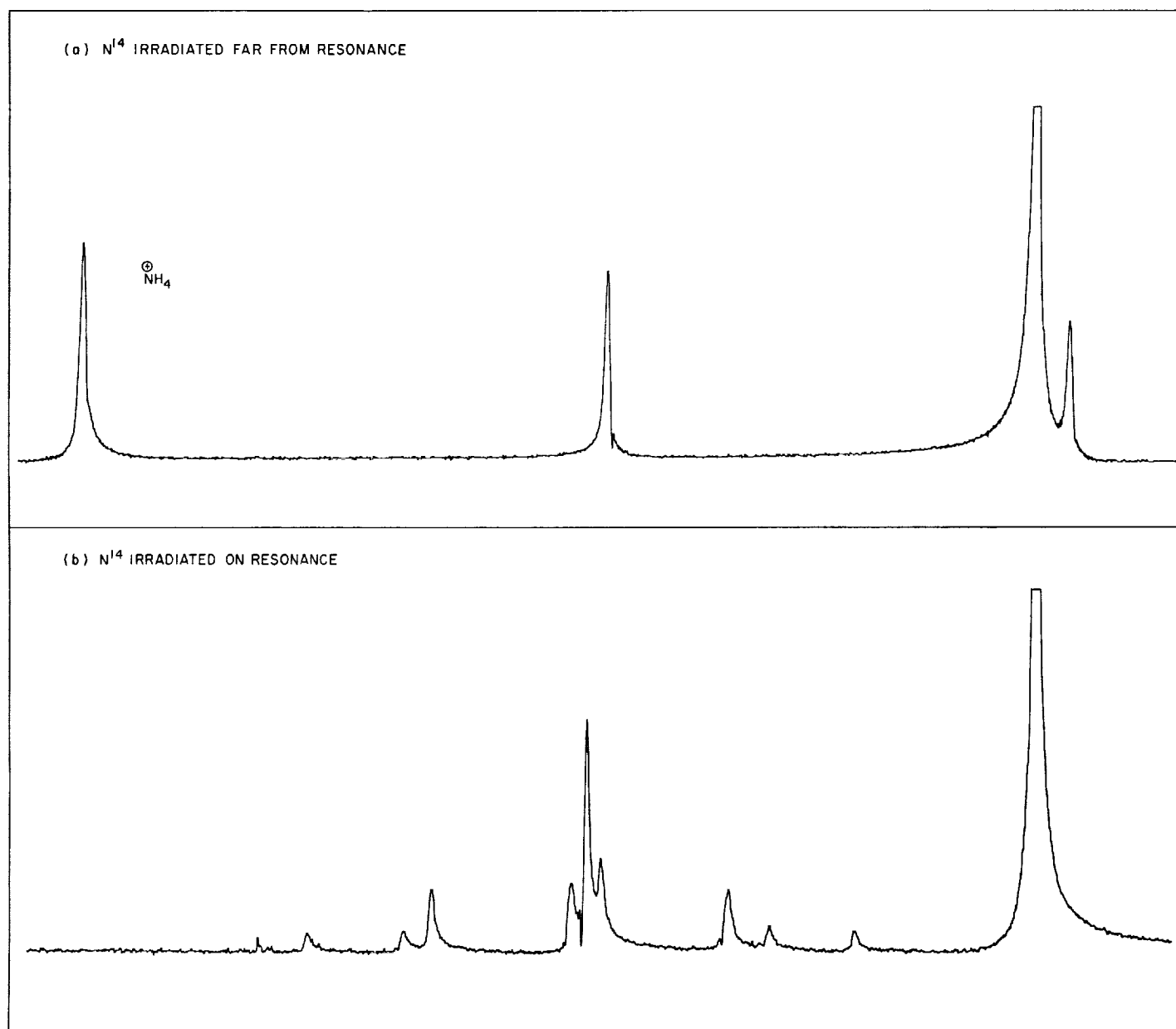
Fig. 4. Spectrum of ammonium ion

(a) NO N^{14} IRRADIATION

(b) N^{14} DECOUPLED FROM PROTONS
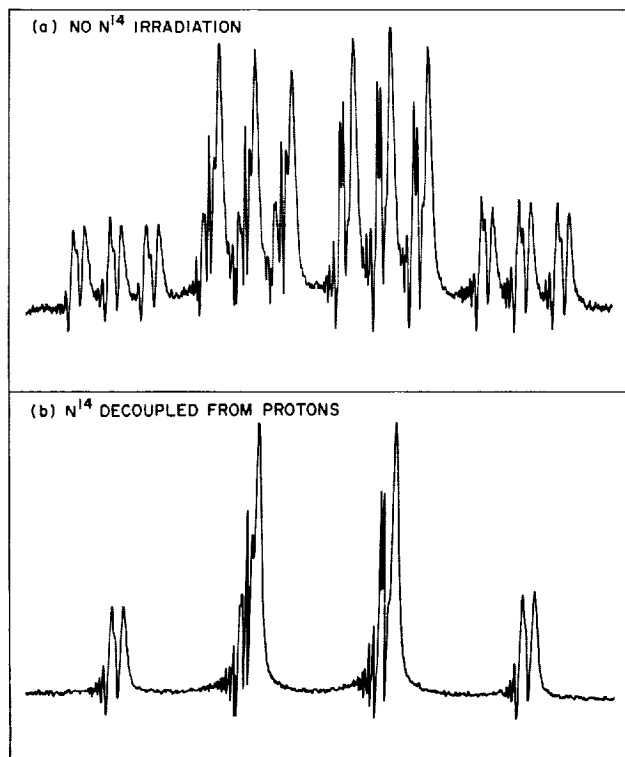
Fig. 5. Spectrum of the CH$_2$-group of ethylisonitrile

B. Some Comments on the Nuclear Magnetic Resonance Spectrum of the ABCD$_3$ System

S. L. Manatt

Only recently have spectra of the ABCD$_3$ type been completely analyzed (Refs. 2 and 3)[1]. There are four chemical-shifted nuclei, and each member of type D has the same spin-spin couplings to A, B, or C by virtue of rapid internal rotation about a carbon–carbon single bond.

The basis spin functions for the ABCD$_3$ nuclear spin problem have not previously been listed. Since they are helpful in understanding some features of the ABCD$_3$ system as well as the double resonance experiments on the system, these functions are presented in Table 1.

[1] Also, DeWolf, M. Y., and Baldeschwieler, J. D. (Private communication).

Considerable factoring of the ABCD$_3$ secular equations results immediately, because there is no mixing between functions of different total z components. Use of permutation symmetry for three magnetically equivalent D nuclei (rotation group C_3) and the appropriate linear combination basis functions results in further factoring and division of the energy levels into one symmetric group and two degenerate antisymmetric groups. The ordering of the basis functions of Table 1 is such that the diagonal matrix elements would be derived in the order of decreasing energy in the case of weak coupling if $\delta_A < \delta_B < \delta_C < \delta_D$, where δ_i is the usual shielding parameter. As seen from Table 1, the diagonalization of the secular equation for the ABCD$_3$ system involves the solution of six 4×4, two 6×6, two 7×7, and one 8×8 eigenvalue problems. Use of the ABCX$_3$ approximation reduces the problem to the solution of twelve 3×3 eigenvalue problems.

The only absorption transitions allowed are those for which $\Delta F_z = 1$ and in which both the initial and final states have the same symmetry. In their analysis of the ABC$_3$ system, Fessenden and Waugh (Ref. 4) have shown transitions between pairs of energy levels from different sets of antisymmetric energy levels (see Table 3 in Ref. 4). The intensity of an absorption transition (i.e., $\beta \to \alpha$ where $\beta = -\frac{1}{2}$ = the lower energy state and $\alpha = \frac{1}{2}$) is proportional to:

$$< \psi_i \mid I^+ \mid \psi_j >^2 ,$$

where ψ_i and ψ_j are the upper and lower eigenfunctions, respectively, and I^+ is the spin raising operator. (The most simple description of the use of I^+ is given in Ref. 5.) In the limit of weak coupling, ψ_i and ψ_j can be replaced by the appropriate basis functions and an estimate of the line intensities obtained; many times this is helpful in making the spectral assignment in the case of intermediate coupling. Using these functions for the ABC$_3$ and ABCD$_3$ systems, it is easy to show that there are no cross transitions such as $a_{-2} \to 1a_{-1}$ or $a_{-2} \to 1a_{-1}$ (see Table 1) between the two groups of antisymmetric energy levels, as indicated by Fessenden and Waugh (Ref. 4). With the inclusion of the appropriate eigenvector components and a little more algebra, it can be shown that this is also true in the strongly coupled case! In this case, it is possible to obtain the intensities of the combination transitions in each separate symmetry group as functions of certain eigenvector components.

By considering all possible permitted transitions between energy levels for the ABCD$_3$ system, one finds that

Table 1. $ABCD_3$ basis functions

	Symmetric				Antisymmetric						
	ABC	DDD	F_z		ABC	DDD	F_z		ABC	DDD	F_z
S_3	ααα	ααα	$+3$								
$1S_2$	$3^{-1/2}$ ααα	$(ααβ + αβα + βαα)$		a_2	$2^{-1/2}$ ααα	$(ααβ - αβα)$	$+2$	a_2	$6^{-1/2}$ ααα	$(ααβ + αβα - 2βαα)$	$+2$
$2S_2$	ααβ	ααα	4×4								
$3S_2$	αβα	ααα	$+2$								
$4S_2$	βαα	ααα									
$1S_1$	$3^{-1/2}$ ααα	$(ββα + βαβ + αββ)$		$1a_1$	$2^{-1/2}$ ααα	$(ββα - βαβ)$		$1a_1$	$6^{-1/2}$ ααα	$(ββα + βαβ - 2αββ)$	
$2S_1$	$3^{-1/2}$ ααβ	$(ααβ + αβα + βαα)$		$2a_1$	$2^{-1/2}$ ααβ	$(ααβ - αβα)$		$2a_1$	$6^{-1/2}$ ααβ	$(ααβ + αβα - 2βαα)$	
$3S_1$	$3^{-1/2}$ αβα	$(ααβ + αβα + βαα)$		$3a_1$	$2^{-1/2}$ αβα	$(ααβ - αβα)$	4×4	$3a_1$	$6^{-1/2}$ αβα	$(ααβ + αβα - 2βαα)$	4×4
$4S_1$	$3^{-1/2}$ βαα	$(ααβ + αβα + βαα)$	7×7	$4a_1$	$2^{-1/2}$ βαα	$(ααβ - αβα)$	$+1$	$4a_1$	$6^{-1/2}$ βαα	$(ααβ + αβα - 2βαα)$	$+1$
$5S_1$	αββ	ααα	$+1$								
$6S_1$	βαβ	ααα									
$7S_1$	ββα	ααα									
$1S_0$	ααα	βββ		$1a_0$	$2^{-1/2}$ ααβ	$(ββα - βαβ)$		$1a_0$	$6^{-1/2}$ ααβ	$(ββα + βαβ - 2αββ)$	
$2S_0$	$3^{-1/2}$ ααβ	$(ββα + βαβ + αββ)$		$2a_0$	$2^{-1/2}$ αβα	$(ββα - βαβ)$		$2a_0$	$6^{-1/2}$ αβα	$(ββα + βαβ - 2αββ)$	
$3S_0$	$3^{-1/2}$ αβα	$(ββα + βαβ + αββ)$		$3a_0$	$2^{-1/2}$ βαα	$(ββα - βαβ)$		$3a_0$	$6^{-1/2}$ βαα	$(ββα + βαβ - 2αββ)$	
$4S_0$	$3^{-1/2}$ βαα	$(ββα + βαβ + αββ)$		$4a_0$	$2^{-1/2}$ αββ	$(ααβ - αβα)$	6×6	$4a_0$	$6^{-1/2}$ αββ	$(ααβ + αβα - 2βαα)$	6×6
$5S_0$	$3^{-1/2}$ αββ	$(ααβ + αβα + βαα)$	8×8	$5a_0$	$2^{-1/2}$ βαβ	$(ααβ - αβα)$	0	$5a_0$	$6^{-1/2}$ βαβ	$(ααβ + αβα - 2βαα)$	0
$6S_0$	$3^{-1/2}$ βαβ	$(ααβ + αβα + βαα)$	0	$6a_0$	$2^{-1/2}$ ββα	$(ααβ - αβα)$		$6a_0$	$6^{-1/2}$ ββα	$(ααβ + αβα - 2βαα)$	
$7S_0$	$3^{-1/2}$ ββα	$(ααβ + αβα + βαα)$									
$8S_0$	βββ	ααα									
$1S_{-1}$	ααβ	βββ									
$2S_{-1}$	αβα	βββ									
$3S_{-1}$	βαα	βββ									
$4S_{-1}$	$3^{-1/2}$ αββ	$(ββα + βαβ + αββ)$	7×7	$1a_{-1}$	$2^{-1/2}$ αββ	$(ββα - βαβ)$		$1a_{-1}$	$6^{-1/2}$ αββ	$(ββα + βαβ - 2αββ)$	
$5S_{-1}$	$3^{-1/2}$ βαβ	$(ββα + βαβ + αββ)$	-1	$2a_{-1}$	$2^{-1/2}$ βαβ	$(ββα - βαβ)$		$2a_{-1}$	$6^{-1/2}$ βαβ	$(ββα + βαβ - 2αββ)$	
$6S_{-1}$	$3^{-1/2}$ ββα	$(ββα + βαβ + αββ)$		$3a_{-1}$	$2^{-1/2}$ ββα	$(ββα - βαβ)$	4×4	$3a_{-1}$	$6^{-1/2}$ ββα	$(ββα + βαβ - 2αββ)$	4×4
$7S_{-1}$	$3^{-1/2}$ βββ	$(ααβ + αβα + βαα)$		$4a_{-1}$	$2^{-1/2}$ βββ	$(ααβ - αβα)$	-1	$4a_{-1}$	$6^{-1/2}$ βββ	$(ααβ + αβα - 2βαα)$	-1
$1S_{-2}$	αββ	βββ									
$2S_{-2}$	βαβ	βββ	4×4								
$3S_{-2}$	ββα	βββ	-2								
$4S_{-2}$	$3^{-1/2}$ βββ	$(ββα + βαβ + αββ)$		a_{-2}	$2^{-1/2}$ βββ	$(ββα - βαβ)$	-2	a_{-2}	$6^{-1/2}$ βββ	$(ββα + βαβ - 2αββ)$	-2
S_{-3}	βββ	βββ	-3								

there are 16 symmetrical transitions and 16 antisymmetrical transitions in 8 degenerate pairs for each of the nuclei A, B, and C. There are 24 symmetrical transitions and 16 antisymmetrical transitions (8 degenerate pairs) for the D_3 nuclei. The remaining transitions are combination transitions involving the simultaneous turning of more than one nuclear spin. These can be divided into two general types: those due to the turning of three spins, such as $5s_0 \rightarrow 1s_{-1}$ (86 symmetrical and 48 antisymmetrical); and those due to the turning of five spins, such as $7s_0 \rightarrow 1s_{-1}$ (12 symmetrical). All the combination transitions should be weak.

The results presented here will subsequently be used in the analysis of the nuclear magnetic resonance spectra of several $ABCD_3$ nuclear spin systems.

C. The Determination of Benzidine Coupled to Carboxymethylcellulose

H. Weetall and N. Weliky

We previously reported a new method for synthesizing an arylaminocellulose by coupling benzidine to carboxymethylcellulose (CMC) in the presence of dicyclohexyl-carbodiimide (Ref. 6). The most direct method for determining the number of benzidine groups coupled is the titration of the acid form of the derivative with NaOH.

A sample of the benzidine derivative of CMC (CMCB) was acidified in 4N HCl, washed with acetone, and titrated with NaOH. Only one inflection point could be found. Apparently the pK of the amino group is close to that of the polycarboxylic acid.

To block the amino groups, 2,4-dinitrophenylsulfonic acid (DNPS) was reacted with the CMCB. The products were acidified and titrated with NaOH. No difference between the treated and untreated CMCB could be found. Although the product was highly colored, coupling must have occurred preferentially with the hydroxyl groups of the cellulose. A similar experiment employing 2,4-dinitrofluorobenzene as a blocking agent gave results similar to those for DNPS. No difference

between the reacted and unreacted CMCB could be detected.

Blocking of the amino groups occurred when benzenesulfonylchloride in tetrahydrofuran was reacted with the derivative (Ref. 7). Titration of the treated material with NaOH showed that 73 to 75% of the carboxyl groups coupled (Table 2).

Table 2. Titration of CMCB treated with benzenesulfonylchloride

CMC degree of substitution, meq/g	CMCB unreacted carboxyl groups, meq/g	Benzidine on CMC, meq/g	Benzidine on CMC, mg/g
0.78	0.21	0.57	104.9
0.78	0.19	0.59	108.6

CMCB diazotized and coupled to protein has been used to isolate rabbit antibody (Ref. 6). We are now comparing this arylaminocellulose with a commercially available form (Cellex-PAB, Bio-Rad, Inc.) regarding its use as an immunoadsorbent, an ion exchange cellulose, and an intermediate in the synthesis of insoluble biologically active materials.

D. The Synthesis and Continual Operation of a Carboxymethylcellulose– Enzyme Column

H. Weetall and N. Weliky

The mechanism of enzyme action is of great interest to chemists and biologists. Many techniques have been developed for studying enzymatic reactions. One which has hardly been explored (Refs. 8–11) is the use of enzyme columns (i.e., enzymes bound to insoluble substances) for studies of kinetics and enzyme–substrate complex formation, and even for the isolation of specific substrates and inhibitors. Also unexplored is the use of enzyme columns for continuous synthesis or degradation of organic compounds. Enzyme columns may also play an important

role in the detection of extraterrestrial life, because enzymatic studies may be carried out without contaminating the samples being tested with soluble foreign protein.

We have been successful in synthesizing a cellulose–enzyme derivative by coupling carboxymethylcellulose (CMC) to the esterase, alkaline phosphatase, through an amide linkage. The feasibility of continuous use of these derivatives, saving large quantities of enzyme, was shown by using a small column continually over a period of 200 hr.

A small Wintrobe disposable pipet was filled with the cellulose–enzyme derivative to a length of 2 cm. The 0.5- × 2.0-cm column was attached to a small LKB Minipump (Stockholm, Sweden), which pumped the substrate through the enzyme column at a constant rate. The product emerging from the column was followed using a spectrophotometer with a continuous-flow 1-mm cell at 410 mμ (Fig. 6).

The substrate used was p-nitrophenylphosphate at a concentration of 0.001M in 1.0M tris buffer, pH 8.0. Upon complete conversion to p-nitrophenol and phosphate ion in a static system, the maximum absorbance was found to be 0.93 at 410 mμ; the phosphate ions do not absorb at this wavelength.

The flow rate through the enzyme column was set at 11 ml/hr so that the optical density recorded was less than that for complete conversion. In this manner we could observe any changes in reaction rate occurring during the experiment. Over the 8-day period that the column was operated, the absorbance of the product was continually recorded. The percent conversion of substrate ($A_{observed}/A_{maximum}$) is plotted versus time in Fig. 7.
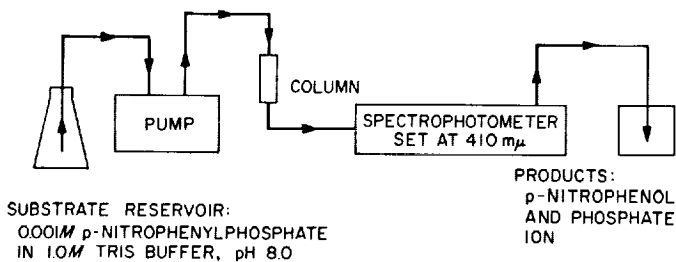
Fig. 6. Instrumentation used in the hydrolysis of p-nitrophenylphosphate by alkaline phosphatase in a packed column
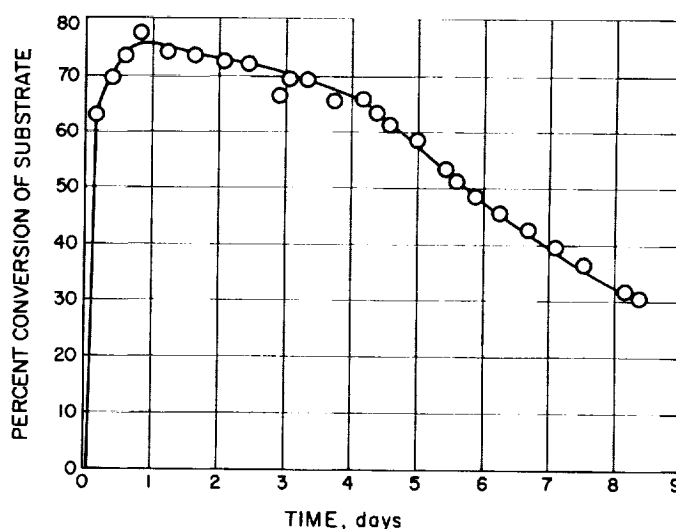
Fig. 7. Percent conversion of p-nitrophenylphosphate vs time

Some discoloration of the column was observed. This may have been caused by oxidation of the substrate or products, denaturation of the protein at room temperature, or other factors. The column was operated uncovered and at room temperature, without precautions for bacterial contamination.

We have shown that a CMC–enzyme column may be synthesized and operated over an extended period of time; similar experiments with other enzymes are now in progress.

E. Additional Results on Metal Plating of Teflon Sheet

S. P. Vango and J. B. Krasinsky

The metal plating of Teflon sheet, carried out in connection with the advanced liquid-propulsion program, has been previously reported (Refs. 12 and 13). Results of additional testing and of some new techniques that have been tried will be presented here.

Some recently lead-plated Teflon sheet has been tested for bond strength, using a Baldwin Lima Hamilton Mark G tensile tester with a crosshead rate of 0.14 in./min.

A 0.0026-in.-thick lead plate was applied to a 10-mil skived Teflon sheet. Four samples (A, B, C, and D) gave bond strengths of 1950, 2590, 1800, and 2430 lb/in², respectively. The range of these samples is approximately equal to the strength of Teflon itself, which ranges from 1200 to 3100 lb/in², depending on the pressures used to compact the powdered Teflon, as well as the sintering temperatures and rate of subsequent cooling (Ref. 14).

Fig. 8 shows the specimens after being tested. Sample B required a pull of 2590 lb/in² for the partial separation of the metal plate from the Teflon. The epoxy bond between the lead surface and holder was broken, but the failure of the metal plate was confined to a very small area. Where the metal plate was pulled from the Teflon, the metal surface still had Teflon adhering to it, indicating that the bond between the metal plate and Teflon was stronger than the Teflon itself.

Fig. 9 shows the long-term permeability test setup. The permeability cells shown are a metal version of those described in Ref. 15. The solution used in the absorption vessel and found to be sufficiently stable consisted of 10 ml of H_2O, 1 ml of 10% H_2SO_4 (vol %), and 0.5 ml of 0.5% metaphenyldiamine hydrochloride (wt %). When some nitrogen tetroxide has been absorbed, the solution changes to light tan due to the formation of nitrous acid. The middle unit in Fig. 9 had a permeable disk; the change in this unit can be seen by contrast with
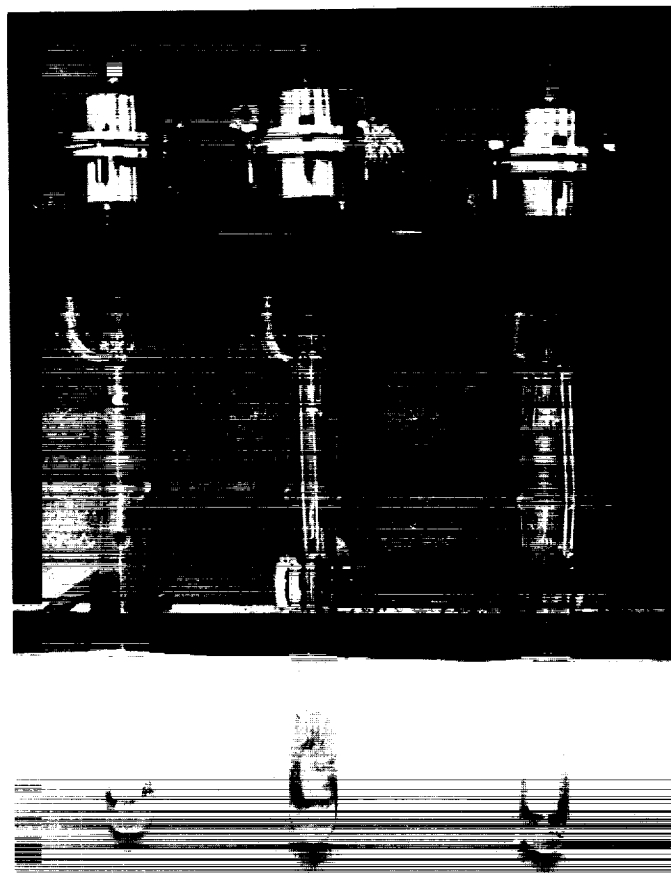
Fig. 8. Bond-strength specimens after testing
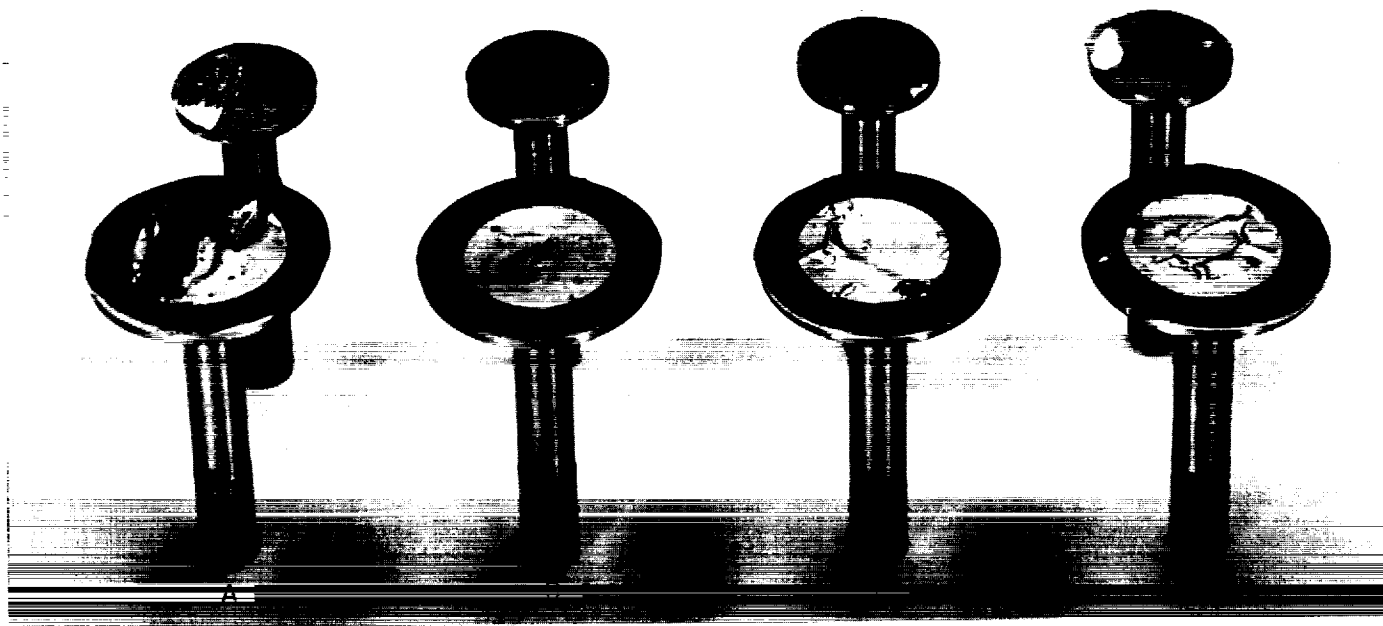
Fig. 9. Long-term permeability test setup

the other two units and also by the discoloration at the fritted surface of the gas dispersion tube.

One lead-plated Teflon sheet (10 mil, skived) has been tested for 241 days without evidence of any permeation of nitrogen tetroxide. A second test has been in progress for 106 days with no permeation resulting thus far.

Ball burnishing of a lead-plated Teflon surface has been tried because of the possible use for this technique in bladder manufacture. After each plating operation (total of three), the Teflon sheet was stapled to a rubber liner in a quart paint can; some solution containing a proprietary burnishing compound was added, as well as some $3/16$-in. stainless-steel balls. The can was then rotated for several hours. The bond strength of the ball-burnished lead plate was 1880 lb/in^2, but the first sample tested for permeability had a slight permeation of nitrogen tetroxide. These tests will be repeated with samples sub-

jected to ball burnishing for about 16 hr after the deposition of each lead coat.

The brush technique for plating has been successfully applied with copper, silver, and gold. These metals have been applied directly to the silver deposit on the Teflon and also to the lead plate. If the gold plate could be applied in a nonporous coat, greater corrosion resistance would be realized; this possibility will, of course, be investigated.

Metal-plated Teflon has many other possible uses; e.g., it can be used for diaphragms, high-temperature circuit boards, and electronic components such as condensers. In addition, Teflon heat shields could be plated to give the desired emissivity, Teflon tubing could be internally plated for use in shielded cables or flexible waveguides, and Teflon with a metal plate could be used for temperature control by resistance heating of the plated metal.

References

1. Elleman, D. D., Pearce, C. D., and Manatt, S. L., "A Control Loop for Regulation of the Magnetic Field and Frequency of a Nuclear Magnetic Resonance Spectrometer," SPS 37-21, Vol. IV, pp. 266–268, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.

2. Bothner-By, A. A., and Naar-Colin, C., Journal of the American Chemical Society, Vol. 83, p. 231, 1961.

3. Finegold, H., Proceedings of the Chemical Society, p. 213, 1962.

4. Fessenden, R. W., and Waugh, J. S., Journal of Chemical Physics, Vol. 30, p. 944, 1959.

5. Roberts, J. D., An Introduction to Spin-Spin Splitting in High-Resolution Nuclear Magnetic Resonance Spectra, pp. 38–40, W. A. Benjamin, Inc., New York, 1961.

6. Weetall, H., and Weliky, N., "A New Method for Synthesizing Arylaminocelluloses," SPS 37-23, Vol. IV, pp. 291–293, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

7. Kamm, O., Qualitative Organic Analysis, p. 69, John Wiley and Sons, Inc., New York, 1948.

References (Cont'd)

8. Grabhofer, N., and Schleith, L., *Naturwissenschaften*, Vol. 40, p. 508, 1953.

9. Barnett, L. B., and Bull, H. B., *Biochimica et Biophysica Acta*, Vol. 36, p. 244, 1959.

10. Bar-Eli, A., and Katchalski, E., *Nature*, Vol. 188, p. 856, 1960.

11. Cebra, J. J., Girol, D., Silman, H. I., and Katchalski, E., *The Journal of Biological Chemistry*, Vol. 236, p. 1720, 1961.

12. Vango, S. P., and Krasinsky, J. B., "Metal Plating of Teflon Sheet as a Means of Reducing Permeation," *SPS 37-21*, Vol. IV, pp. 251, 252, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.

13. Vango, S. P., and Krasinsky, J. B., "Continued Effort in Metal Plating of Teflon Sheet as a Means of Reducing Permeation," *SPS 37-23*, Vol. IV, pp. 283–286, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

14. *Teflon*, available from E. I. DuPont De Nemours & Co., Wilmington, Delaware.

15. Vango, S. P., *Determination of Permeability of Teflon Sheet to Nitrogen Tetroxide and Hydrazine*, Technical Memorandum No. 33-55, Jet Propulsion Laboratory, Pasadena, California, August 1961.

XV. Fluid Physics

A. The Inviscid Stability of the Compressible Laminar Boundary Layer: Part II

L. M. Mack

A numerical method was presented in Ref. 1 for the computation of the inviscid stability characteristics of the compressible laminar boundary layer on a flat plate. Several results were presented, including a curve of α, the dimensionless wave number of the neutral inviscid subsonic disturbance, as a function of M_1, the free-stream Mach number. It was also noted that, for sufficiently large Mach numbers, the wave number is no longer single-valued, and that, for $M_1 > 4.5$, the upper branch of the neutral stability curve approaches the second value of α, rather than the first, as the Reynolds number approaches infinity.

It has now been established that α is not only double-valued, but is multiple-valued at each Mach number.

There is, of course, only a single value of the phase velocity, c_r, which is fixed by the requirement that an inviscid subsonic neutral disturbance travel with a speed equal to the mean boundary-layer velocity at the point where the derivative of w'/T is zero, where w' is the derivative of the mean velocity and T is the mean temperature. Fig. 1 gives the results obtained to date up to $M_1 = 10$. The lower curve is the same as was given in Ref. 1. At $M_1 = 10$, ten values of α were obtained. Starting at about the fifth value, the increment in α from one curve to the other is nearly constant. There is no reason to believe that there is any upper limit on the possible value of α, and presumably an infinite number of separate solutions exist.

The shapes of the curves in Fig. 1 point to the existence of an asymptote somewhere between $M_1 = 2$ and 3. For $M_1 < 2$, no multiple solutions could be found, a situation in accord with the analytic proof that α is unique as long as there is no region in the boundary layer in which the mean flow relative to the disturbance wave is supersonic. A supersonic relative region first appears in the flat-plate boundary layer at approximately $M_1 = 2.2$ and is present
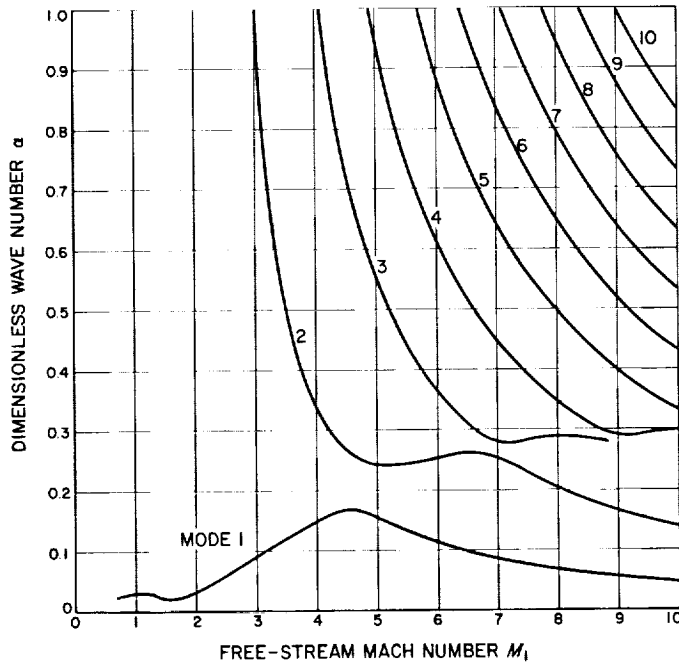
Fig. 1. Wave numbers of the first ten modes as functions of Mach number (neutral inviscid subsonic disturbances)

at all higher Mach numbers. It was the impossibility of finding a uniqueness proof for the case of a supersonic relative region that led Lees and Reshotko (Ref. 2) to conjecture that possibly α is non-unique under these conditions.

The nature of the multiple solutions can be clarified by examining the flow relative to the disturbance wave. In this coordinate system the flow is steady with sinusoidal streamlines. The Mach number of the relative flow is

$$M = \frac{u}{T^{1/2}} M_1 , \tag{1}$$

where $u = c_r - w$ is the mean velocity relative to the wave. As shown in Ref. 3, the pressure–fluctuation amplitude function π can be written

$$\frac{\pi}{\gamma M^2} = i \frac{1}{1 - M^2} \left(\frac{\phi}{u} \right)' , \tag{2}$$

where γ is the ratio of specific heats. The prime refers to differentiation with respect to the boundary-layer coordinate, and ϕ, when multiplied by α, is the amplitude function of the normal disturbance velocity. Therefore, ϕ/u is proportional to the amplitude function of the streamline slope. Eq. (2) can be recognized as a form of the linearized pressure–area relation of one-dimensional

isentropic flow, and $(\phi/u)'$ is identified as the amplitude function of the streamtube area change. The latter function must be zero at $M = 1$, the sonic point of the relative flow, since π is continuous through the sonic point. If $(\phi/u)'$ is zero at any other point in the boundary layer, π will also be zero and will undergo a sign change which signifies a 180-deg phase shift in the pressure fluctuation.

The distribution of ϕ/u across the boundary layer has been computed for the first three values of α at $M_1 = 5.8$ (Fig. 2). For a neutral disturbance with the normalization used here, ϕ is pure imaginary; thus, the quantity plotted
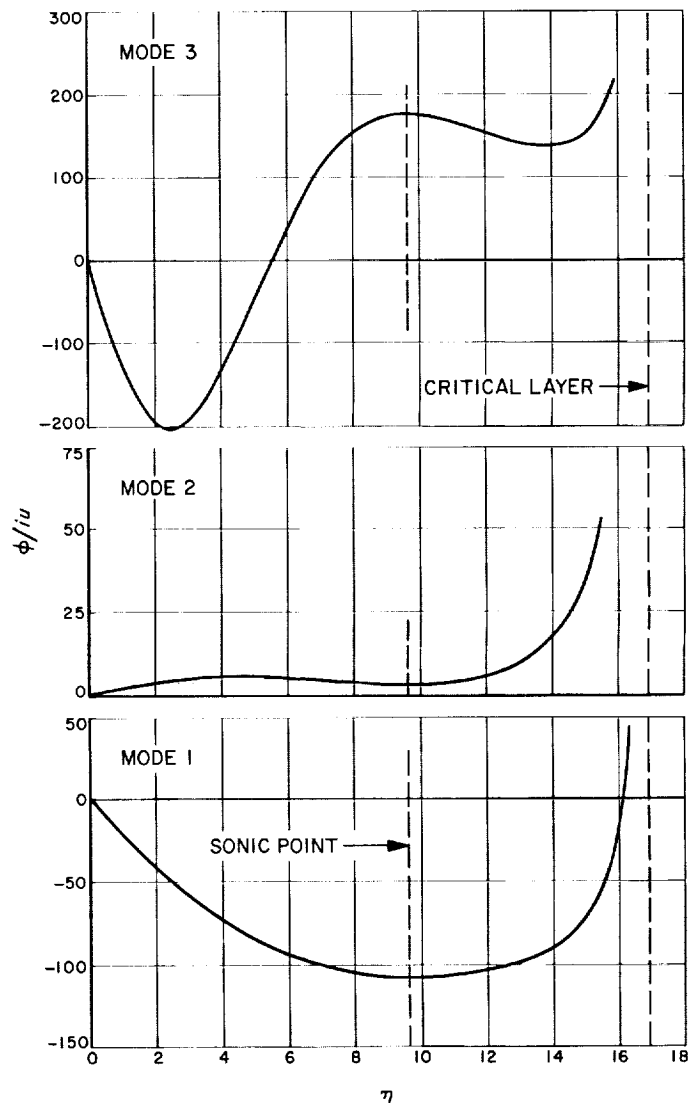
Fig. 2. Distribution through the boundary layer of streamline-slope amplitude functions at $M_1 = 5.8$ for the first three modes (neutral inviscid subsonic disturbances)

in Fig. 2 is ϕ/iu. Only the region below the critical layer ($w = c_r$) is shown. Above the critical layer all distributions are similar in shape. Inspection of Fig. 2 reveals the multiple solutions to be a sequence of vibration modes of the boundary layer. The first mode, corresponding to the lowest value of α, has a single extremum of ϕ/u. This extremum is located at the sonic point of the relative flow, as it must be. For this mode the pressure fluctuation has the same sign throughout the boundary layer. The second mode has two extrema: one at the sonic point and the other in the supersonic region. Consequently, π is zero at the second extremum and there is a 180-deg phase shift in the pressure fluctuation. The third mode has three extrema (the third being in the subsonic region) and two zeros in the pressure–fluctuation amplitude

function. This pattern continues for higher values of α. For instance, the tenth value of α has ten extrema and nine 180-deg phase shifts in the pressure fluctuation.

Another feature of the multiple solutions is the rapid increase in amplitude of the higher modes. (The normalization is such that the magnitude of the longitudinal velocity fluctuation is always 1.0 at the edge of the boundary layer.) The large peak values of ϕ/u and the large number of extrema lead to correspondingly large magnitudes of π. Eventually the magnitudes become so large ($\sim 10^7$) that the numerical procedure loses accuracy and can no longer satisfy the boundary condition of zero normal velocity at the wall. As a result, further modes cannot be obtained.

References

1. Mack, L. M., "The Inviscid Stability of the Compressible Laminar Boundary Layer," SPS 37-23, Vol. IV, pp. 297–301, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

2. Lees, L., and Reshotko, E., "Stability of the Compressible Laminar Boundary Layer," Journal of Fluid Mechanics, Vol. 12, Part 4, pp. 555–590, April 1962.

3. Mack, L. M., Numerical Calculation of the Stability of the Compressible Laminar Boundary Layer, Report No. 20-122, Jet Propulsion Laboratory, Pasadena, California, May 1960.

XVI. Physics

A. Efficiency of Propellant Heating in Vortex Tubes

H. J. Stumpf

The ultimate performance potential of a gaseous-vortex propulsion reactor is dominated by its internal thermal radiation. A simplified calculation of the thermal radiation heat flux at a vortex tube wall has been carried out in order to estimate the fraction of the total fission power released in the cavity regions utilized to heat the propellant. The simplifying assumptions are listed in Ref. 1.[1] The results of this calculation are given in Fig. 1. It can be seen that the heat flux at the wall increases with the tube radius r_t and the temperature ratio $T_c/T_w = I^2$, where T_c is the central core temperature and T_w is the wall temperature. At constant values of I^2, the heat flux at first increases rapidly with r_t and then increases more slowly since it approaches the case of an infinite flat plate

---

[1] Also, Stumpf, H. J., *Heat Loss from Vortex Tubes* (To be published).

receiving radiant energy from an infinite gas volume. At constant values of r_t, the heat flux increases rapidly for $I^2 > 3.75$, since the gas in the tube then consists of a central core radiating as a blackbody surrounded by a blanket of relatively transparent gas. This is actually a more pessimistic case than that of the uniformly black or uniformly transparent gas analyzed previously in Ref. 2.

It is interesting to note that contemporary regeneratively cooled nozzles for chemical engine applications boast a surface heat flux of 20 Btu/sec-in.^{2} = 3.27 kw/cm^2. From Fig. 1 it is apparent that, for temperature ratios > 5, the surface heat flux exceeds this value and may result in wall cooling problems.

A power balance to determine the heat input to the propellant was carried out utilizing the results of Ref. 3.[1] The fraction of the total cavity power used to heat the propellant is shown in Fig. 2. For temperature ratios ≤ 4, an appreciable fraction of the cavity power appears

Fig. 1. Thermal radiation heat flux at vortex tube wall as a function of vortex tube radius

Fig. 2. Fraction of total power released in vortex tubes utilized to heat propellant as a function of vortex tube radius

as propellant heating. For temperature ratios > 5, however, less than two-thirds of the power is utilized in heating the propellant even at small tube radii, and this fraction decreases for the larger tube radii.

It should be stressed that these results are based upon an assumed pressure of 100 atm in the vortex tube. At this operating point, it appears that large radii and high exhaust temperatures lead to excessive radiation losses and an inefficient propellant heating process. It would be advantageous to use higher system pressures since the optical opacity of the gas would be higher, thus screening the wall from some of the radiation from the high-temperature central region.

B. Rocket Measurement of the Nitric Oxide Dayglow

C. A. Barth

A scanning ultraviolet spectrometer was flown from Wallops Island, Virginia, on a NASA Aerobee 4.85 on

November 17, 1963, at 1:15 pm EST to an altitude of 184 km. The 0.5-m spectrometer scanned the dayglow spectrum from 1500 to 3200 A. The rocket's attitude-control system aimed the spectrometer at the zenith during the ascent and at the horizon during the descent. Spectral scans of the ultraviolet dayglow were obtained in the zenith direction in the altitude range from 62 to 110 km, and also (with greatly increased intensity) during the scan of the horizon. The most prominent features of these dayglow spectral scans were the gamma bands of nitric oxide.

The spectrometer was a 0.5-m Fastie–Ebert instrument (Ref. 4). The grating had a 102- × 102-mm ruled area, with 2160 lines/mm, and was blazed at a nominal 2000 A. A single 2-in. photomultiplier with a cesium telluride photocathode and a sapphire window (ASCOP Type 543F-05-14) was used as the detector. The instrument had its maximum sensitivity at 2150 A and, because of the high-work-function photocathode, the sensitivity dropped 3 orders-of-magnitude in going from 2150 to 3200 A. The spectrum was scanned from 1500 to 3200 A and back in 7.5 sec. The instrument was equipped with programmed slits which allowed the downward scan to be made at 4-A resolution and the upward scan, which was three times faster, at 10-A resolution. The spectrometer was aligned with the longitudinal axis of the rocket, and the entrance slit was shielded from direct sunlight by a sunshade. A wavelength calibration was performed in flight with a mercury lamp in the nose-tip of the rocket.

The rocket was launched when the Sun was at a zenith angle of 60 deg. At an altitude of 55 km, the nose-tip was ejected and the spectral scans of the atmosphere were begun. The rocket despun and the attitude-control system erected the rocket to the vertical. Between 62 and 110 km, six complete spectral scans were obtained with the spectrometer viewing the zenith. Then, an unplanned maneuver by the attitude-control system caused the rocket to begin a slow turnover at about 0.5 deg/sec. This maneuver caused a slow scan of the horizon late in the flight when the rocket was at an altitude of 144 km. Six complete spectral scans were obtained as the spectrometer's zenith angle changed from 80 to 105 deg. The most intense spectral scans were recorded during this portion of the flight.

One of the spectral scans, obtained at 144 km during the horizon scan, is shown in the lower part of Fig. 3. The zenith angle was 100 ± 5 deg, and the field of view was 12 deg. The spectrum presented here extends from
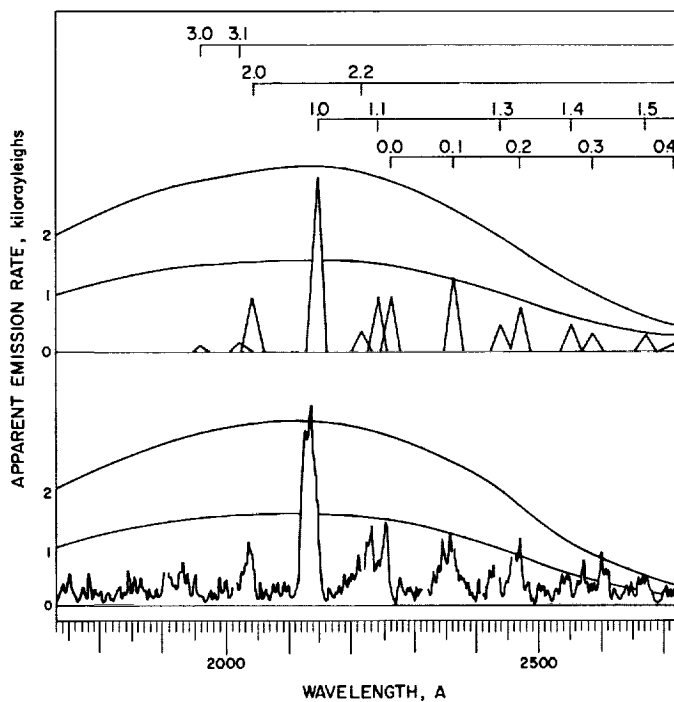
Fig. 3. Nitric oxide dayglow spectrum from 1730 to 2730 A

1730 to 2730 A at a resolution of 10 A. It represents 1 sec of data out of the 360 sec that were obtained during the flight. The spectrum has been traced from the original telemetry record. Six obvious noise pulses have been edited out. The spectral response of the intrument has been sketched on the spectrum, showing the 0-, 1-, and 2-kilorayleigh intensity levels. The cesium telluride photomultiplier causes the sensitivity to drop at longer wavelengths; the grating blaze and sapphire window of the photomultiplier caused the fall-off toward shorter wavelengths.

The flight spectrum was compared to a synthetic spectrum calculated in the following way: For single scattering, the fluorescent photon emission rate at height z is given by (Ref. 5):

$$4\pi \, \mathscr{J} \, (\mu) = \frac{g \mathfrak{N}(z)}{\mu},\qquad(1)$$

where $\mu = \cos \theta$, θ being the zenith angle of the scattered radiation, and $\mathfrak{N}$ is the number of molecules in a cm² column above height z. The number of photons scattered per sec per unit molecule for each spectral band $v'-v''$ is:

$$g_{v'-v''} = \pi \mathscr{F}_{v'-0} \frac{\pi e^2}{mc} f_{v'-0} \frac{A_{v'-v''}}{\sum_{v''} A_{v'-v''}},$$

where $\pi\mathscr{F}_{v'-0}$ is the solar flux averaged over the absorption band (Refs. 6 and 7), $f_{v'-0}$ is the f-value of the absorption band (Ref. 8), and $A_{v'-v''}$ is the transition probability of the emission band (Ref. 8). The quantity $g_{v'-v''}$ was calculated for 12 of the most intense bands of the gamma system of nitric oxide. The physical quantities used in the calculation and the results are given in Table 1. Since Eq. (1) shows that the emission rate is directly proportional to the quantity $g_{v'-v''}$ for any given amount of nitric oxide, the relative intensity distribution of the bands may be obtained from the last column of Table 1. The 1–0 band at 2155 A is the most intense gamma band in fluorescence. The calculated intensities of the dayglow gamma bands when multiplied by the instrument spectral response are shown in the upper part of Fig. 3. The bands are identified according to progression in the upper right-hand portion of Fig. 3. Although the gamma bands appear with double heads at moderate resolution, the study of a laboratory source with the flight instrument showed that, at 10-A resolution, the double heads cannot be distinguished and there is only a hint of the band degradation to shorter wavelengths. A comparison of the synthetic spectrum in the upper part and the flight spectrum in the lower part of Fig. 3 shows that the nitric oxide gamma bands are a part of the ultraviolet dayglow spectrum of the Earth. A detailed comparison indicated that the grating cam drive of the flight spectrometer was nonlinear; the wavelength calibration had been performed at the ends of the spectrum being analyzed. It should be realized that this particular flight spectrum was obtained by moving the grating through 1000 A in 1 sec.

The measurement of the intensity of the 1–0 gamma band at 2155 A permits the amount of nitric oxide present in the upper atmosphere to be determined. A spectrum obtained at 85 km with the instrument pointing in the zenith direction shows the unblended band with an emission rate of 830 rayleighs. Using the quantity g_{1-0} from Table 1, the column density of nitric oxide above the 85-km level is 1.7×10^{14} molecules/cm^2.

Previous attempts to measure neutral nitric oxide molecules in the upper atmosphere have not resulted in the detection of nitric oxide, but have only indicated an upper limit on the abundance. Jursa, Tanaka, and LeBlanc (Ref. 9) attempted to measure nitric oxide by absorption spectroscopy, using a rocket spectrograph with a solar-pointing control. They were unable to place an upper limit of 10^{15} molecules/cm^2 above the altitude range of the experiment (63 to 87 km).

Table 1. Emission rate factors for the dayglow of the nitric oxide gamma bands

Band $v'-v''$	$\lambda_{v'-v''}$, A	$\pi\mathscr{F}_{v'-0}$, photons/cm^2-sec-A $\times 10^{11}$	$f_{v'-0}$, $\times 10^{-3}$	$\dfrac{A_{v'-v''}}{\sum\limits_{v''} A_{v'-v''}}$	$g_{v'-v''}$, photons/sec $\times 10^{-6}$
2–0	2053	1.6	6.73	0.366	1.5
1–0	2155	5.1	7.88	0.292	4.8
2–2	2223	—	—	0.176	7.0
1–1	2246	—	—	0.088	1.4
0–0	2269	6.0	3.99	0.138	1.5
0–1	2370	—	—	0.226	2.4
1–3	2447	—	—	0.063	1.0
0–2	2479	—	—	0.211	2.3
1–4	2559	—	—	0.130	2.1
0–3	2596	—	—	0.158	1.7
1–5	2680	—	—	0.153	2.5
0–4	2722	—	—	0.116	1.3

C. Short-Time Behavior of an Initial Step Discontinuity in a Collisionless Plasma

J. S. Zmuidzinas and Y. Hiroshige

The investigation of the phenomenon of a free expansion into a near vacuum of an initially confined collisionless plasma is of importance for understanding many practical problems such as the operation of an ion engine in space or that of a plasma shock tube; it is also of an intrinsic physical and mathematical interest due to the presence of a large space inhomogeneity.

For a non-ionized gas, similar problems have been studied by a number of authors. Molmud (Ref. 10) has obtained a solution for a gas initially confined in a sphere. Using the collisionless Boltzmann equation, Keller (Ref. 11) found a solution for the one-dimensional expansion of a gas from a half-space. A related problem, studied numerically by Dawson (Ref. 12), is the observation of the effect of a large sinusoidal disturbance on an initially cold plasma.

In the following, we shall derive an integral equation which describes the one-dimensional transient motion of the plasma resulting from the sudden removal of a partition confining the plasma to a half-space. Qualitatively, we should expect plasma oscillations to be excited about

the average motion of the plasma, which is at the ion thermal speed. In formulating the problem, we make the following assumptions:

(1) The flow is one-dimensional. Both upstream and downstream chambers are of infinite length.

(2) Initially, electrons and ions are separately in thermal equilibrium in both the upstream and downstream chambers.

(3) The plasma is collisionless for all times of interest and is described by a set of Vlasov equations.

(4) The forces due to induced magnetic fields are negligible in comparison with those due to induced electric fields.

(5) A linearization of the distribution functions can be made about the initial ones for times during which the wavefront does not advance significantly. The condition of validity is

$$| n(x, t) - n(x, 0) | \ll n(x, 0),$$

where n is either the electron or the ion number density. Of course, it is understood that the temperature differences between the two chambers cannot be arbitrarily large.

(6) The distribution functions vanish at spatial infinity in both directions for all times.

The equations governing the problem are the usual Vlasov set of equations:

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} - \frac{e}{m} E \frac{\partial f}{\partial v} = 0,$$

$$\frac{\partial F}{\partial t} + v \frac{\partial F}{\partial x} + \frac{e}{M} E \frac{\partial F}{\partial v} = 0, \qquad (1)$$

$$\frac{\partial E}{\partial x} = 4\pi e \int_{-\infty}^{\infty} dv \, (F - f),$$

where m and M are the masses of electrons and ions; f and F, their respective distribution functions; E, the electric field; and x, v, and t, the space, velocity, and time variables.

The initial and boundary conditions are the following:

$$f(x, v, t = 0) = g(x, v) = \left[g_1(v) \eta(a - x) + g_2(v) \eta(x - a) \right]$$

$$\times \exp(-\alpha | x - a |),$$

$$f(x = \pm \infty, v, t) = 0,$$

where

$$\alpha \to 0^+,$$

$$g_i(v) = n_i (m/2\pi K\theta_i)^{\frac{1}{2}} \exp(- mv^2/2K\theta_i), \quad i = 1, 2,$$

are the initial Maxwellian distribution functions for electrons at temperatures θ_1 and θ_2 with number densities n_1 and n_2, with K being the Boltzmann constant. The subscripts 1 and 2 denote the upstream and downstream chambers, respectively. The partition is at $x = a$, and η represents the unit step function. A similar set of conditions is assumed for the ion distribution function; we shall not display these since they are obtained by simply replacing appropriate lower-case letters by capitals.

The linearization of Eq. (1) amounts to the replacement $\partial f/\partial v \to \partial g/\partial v$ and $\partial F/\partial v \to \partial G/\partial v$. Introducing the Fourier-Laplace transform pairs

$$\widetilde{\phi}(k, v, \omega) = \int_{-\infty}^{\infty} dx \int_{0}^{\infty} dt \exp\left[-i(kx - \omega t) \right] \phi(x, v, t),$$

$$\phi(x, v, t) = (2\pi)^{-2} \int_{-\infty}^{\infty} dk \int_{-\infty+i\omega_0}^{\infty+i\omega_0} d\omega \exp\left[i(kx - \omega t) \right]$$

$$\times \widetilde{\phi}(k, v, \omega),$$

where ω_0 is a suitable positive constant, we obtain from Eq. (1) the transformed set:

$$\widetilde{f}(k, v, \omega) = \frac{\omega_p^2}{2\pi n_1}$$

$$\times \int \frac{dk'}{k'} \frac{\widetilde{h}(k', \omega)}{\omega - kv} \frac{\partial}{\partial v} \widetilde{g}(k - k', v), \qquad (2)$$

$$F(k, v, \omega) = -\frac{\Omega_p^2}{2\pi n_1}$$

$$\times \int \frac{dk'}{k'} \frac{\widetilde{h}(k', \omega)}{\omega - kv} \frac{\partial}{\partial v} \widetilde{G}(k - k', v). \qquad (3)$$

Here,

$$\widetilde{h}(k, \omega) = \int dv \left[\widetilde{F}(k, v, \omega) - \widetilde{f}(k, v, \omega) \right],$$

$$\widetilde{g}(k, v) = \int dx \exp(-ikx) g(x, v)$$

$$= i \exp(-ika) \left[\frac{g_1(v)}{k + i\alpha} - \frac{g_2(v)}{k - i\alpha} \right],$$

$$\omega_p^2 = 4\pi n_1 e^2/m,$$

and similarly for $\widetilde{G}$ and Ω_p^2. Subtracting Eq. (2) from Eq. (3), we find the following integral equation for $\widetilde{h}$:

$$\widetilde{h}(k, \omega) + \int dk' \, Q(k, k', \omega) \widetilde{h}(k', \omega) = 0, \qquad (4)$$

where

$$Q(k, k', \omega) = \frac{1}{2\pi n_i k} \int dv \, \frac{1}{\omega - kv}$$

$$\times \frac{\partial}{\partial v} \left[\Omega_p^2 \widetilde{G}(k - k', v) + \omega_p^2 \widetilde{g}(k - k', v) \right].$$

This equation mathematically describes our physical problem.

The distribution functions for both electrons and ions are obtained by substituting the solution of the integral equation into Eqs. (2) and (3) and taking the inverse transforms. However, for investigation of the properties of initially induced plasma oscillations, it is not necessary to solve the complete equations. The properties of the integral Eq. (4) and of the dispersion relation will be discussed at a later date.

D. Influence of Electron–Phonon Interaction on the Screening of a Test Charge in a Metal

O. von Roos

Some time ago the dielectric constant of an electron–phonon system was derived within the random phase approximation (Ref. 13). The resulting expression may be written as

$$\epsilon(K, \omega) = 1 - A(K, \omega) - \frac{\omega_i^2}{\omega^2 + \omega_i^2 - \left[\omega_L(K) \right]^2}, \qquad (1)$$

where ω_i is the ionic plasma frequency and $\omega_L(K)$ is the unrenormalized frequency of longitudinal sound waves. Furthermore, A is given by

$$A = \omega_P^2 \int d^3k \, \frac{F_0(k)}{\left(\omega - \frac{\hbar}{m} K \cdot k \right)^2 - \left(\frac{\hbar}{2m} K^2 \right)^2}, \qquad (2)$$

where ω_P is the electronic plasma frequency and F_0 is the Fermi distribution function. Letting $\omega_i = (m/M)\omega_P$ go to zero in Eq. (1) is tantamount to neglecting the electron–

phonon interaction, and Eq. (1) reduces to Lindhard's result (Ref. 14). Now, it is well-known that

$$\left[\omega_L(K) \right]^2 + \omega_1^2 + \omega_2^2 = \omega_i^2, \qquad (3)$$

where ω_1 and ω_2 are the frequencies of the two transverse sound waves (Ref. 15). In the long wavelength limit, the transverse modes are not affected by the electron screening as is the longitudinal sound. Introducing, therefore, the velocities of transverse sound c_1 and c_2, we have from Eq. (3):

$$\left[\omega_L(K) \right]^2 = \omega_i^2 - (c_1^2 + c_2^2) K^2. \qquad (4)$$

The Fourier transform of the electric field of a point charge q at rest in vacuum is:

$$\mathbf{D} = - i \mathbf{K} \frac{4\pi q}{K^2} \delta(\omega). \qquad (5)$$

However, placing this charge into the plasma characterized by the dielectric constant Eq. (1), we have

$$\mathbf{E} = - i \mathbf{K} \frac{4\pi q}{\epsilon(K, 0) K^2} \delta(\omega). \qquad (6)$$

The associated potential in configuration space is then given by:

$$\phi = \frac{4\pi q}{(2\pi)^3} \int d^3K \left[K^2 \epsilon(K, 0) \right]^{-1} e^{i\mathbf{K} \cdot \mathbf{r}}. \qquad (7)$$

For large enough distances r, significant contributions toward the integral Eq. (7) occur only for small K. In this case,

$$\epsilon(K, 0) = 1 + \frac{1}{K^2} \left(\frac{3 \omega_P^2}{v_F^2} - \frac{\omega_i^2}{c_1^2 + c_2^2} \right). \qquad (8)$$

Hence, ϕ from Eq. (7) is given by:

$$\phi = \frac{q}{r} e^{-r/r_0}, \qquad (9)$$

with the screening constant

$$r_0 = \left(\frac{3 \omega_P^2}{v_F^2} - \frac{\omega_i^2}{c_1^2 + c_2^2} \right)^{-1/2}, \qquad (10$$

provided that $r_0 > 0$. From the well-known expression

$$v_L = \left(\frac{m}{3M} \right)^{1/2} v_F \qquad (11)$$

for the longitudinal velocity of sound (Ref. 15), we have from Eq. (10):

$$r_0 = \omega_i \left(\frac{1}{c_L^2} - \frac{1}{c_1^2 + c_2^2} \right)^{-1/2}. \qquad (12)$$

Now, $r_0' = \omega_1/c_L$ is the screening distance without phonons. We obtain, therefore, an increase in screening length due to electron–phonon interactions; namely,

$$\alpha = \frac{r_0}{r_0'} = \left(1 - \frac{c_L^2}{c_1^2 + c_2^2}\right)^{-\frac{1}{2}}. \tag{13}$$

For sodium and potassium, $\alpha = 1.4$ and 1.25, respectively.

E. Relativistic Many-Particle States

J. S. Zmuidzinas

In the course of reducing tensor products of irreducible representations of the inhomogeneous Lorentz group, one again obtains irreducible representations whose states are degenerate with respect to this group. Thus, in addition to the usual "external" quantum numbers specifying the state within each irreducible representation, one also has a set of degeneracy labels or "internal" quantum numbers which indicate how the tensor product was reduced. Here we wish to outline a group-theoretical method of constructing many-particle states characterized by such sets of external and internal quantum numbers.

The Lie algebra of the restricted inhomogeneous Lorentz or the restricted Poincaré group $P_+^{\uparrow}$ is spanned by the ten (Hermitian) generators P_μ and $M_{\mu\nu} = -M_{\nu\mu}$, with $\mu = 0, 1, 2$, and 3, of space-time translations and rotations, respectively. They satisfy the well-known commutation relations:

$$[P_\mu, P_\nu] = 0,$$

$$[M_{\mu\nu}, P_\rho] = i(P_\mu g_{\nu\rho} - P_\nu g_{\mu\rho}),$$

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(M_{\mu\sigma} g_{\nu\rho} - M_{\mu\rho} g_{\nu\sigma} + M_{\nu\rho} g_{\mu\sigma} - M_{\nu\sigma} g_{\mu\rho}), \tag{1}$$

with $g_{00} = -g_{ii} = 1$; $i = 1, 2$, and 3; and $g_{\mu\nu} = 0$ for $\mu \neq \nu$. Introducing the (pseudo) 4-vector

$$W_\mu = \tfrac{1}{2} \epsilon_{\mu\nu\rho\sigma} M^{\nu\rho} P^\sigma, \qquad \epsilon_{0123} = -1,$$

with the commutation relations

$$[P_\mu, W_\nu] = 0,$$

$$[W_\mu, W_\nu] = -i \epsilon_{\mu\nu\rho\sigma} W^\rho P^\sigma,$$

$$[M_{\mu\nu}, W_\rho] = i(W_\mu g_{\nu\rho} - W_\nu g_{\mu\rho}),$$

one finds that the operators

$$C_1 = P_\mu P^\mu,$$

$$C_2 = W_\mu W^\mu$$

commute with all P_μ and $M_{\mu\nu}$ and, hence, are invariants (Casimir operators) of the Lie algebra. Irreducible (unitary) representations of $P_+^{\uparrow}$ are labeled by the eigenvalues m^2 and $s(s + 1)$ of C_1 and $-C_2/m^2$, respectively. Within each irreducible representation (m, s), the states are further distinguished by the eigenvalues of operators belonging to a maximal Abelian subalgebra of the Lie algebra (or rather of its enveloping algebra). One such subalgebra consists of the momentum and helicity operators P_i, $i = 1, 2$, and 3, and $W_0/|\mathbf{P}|$, with eigenvalues $P_i(-\infty < P_i < \infty)$ and $|\lambda| \leq s$ in integer steps. The one-particle states $|m\,s\,\mathbf{p}\,\lambda>$ are thus eigenstates of six commuting operators. A further invariant is the sign of the energy, $\epsilon = p_0/|p_0|$; we shall ignore it by restricting our discussion to the case $\epsilon = +1$.

Many-particle states of free non-interacting particles are obtained by taking tensor products of one-particle states. The n-particle state

$$|m_1 s_1 \mathbf{p}_1 \lambda_1 > \otimes \cdots \otimes | m_n s_n \mathbf{p}_n \lambda_n >$$

is an eigenstate of $6n$ commuting operators, namely, $C_1^{(k)}$, $C_2^{(k)}$, $\mathbf{P}^{(k)}$, and $W_0^{(k)}/|\mathbf{P}^{(k)}|$, with $k = 1, 2, \cdots, n$. We shall ultimately be interested in writing this state as a linear combination of eigenstates with certain external and internal quantum numbers; to proceed with the construction of the latter states, let us first consider the case $n = 2$.

A two-particle system can be characterized by the state of its center of mass and by the configuration of the two particles in their center-of-mass frame. We define the external or center-of-mass generators:

$$P_\mu = P_\mu^{(1)} + P_\mu^{(2)},$$

$$M_{\mu\nu} = M_{\mu\nu}^{(1)} + M_{\mu\nu}^{(2)},$$

where each of the sets $\{P_\mu^{(1)}, M_{\mu\nu}^{(1)}\}$ and $\{P_\mu^{(2)}, M_{\mu\nu}^{(2)}\}$ satisfy Eq. (1) and $[X^{(i)}, X^{(j)}] = 0$, with $i \neq j$ and X being any generator in the Lie algebra of $P_+^{\uparrow}$. It follows that $P_\mu, M_{\mu\nu}$ also satisfy Eq. (1); therefore, one may take $C_1, C_2, \mathbf{P}$, and $W_0/|\mathbf{P}|$ as the set of commuting external operators for purposes of labeling the two-particle states. Note that

$$W_\mu = \tfrac{1}{2} \epsilon_{\mu\nu\rho\sigma} M^{\nu\rho} P^\sigma$$

$$= \tfrac{1}{2} \epsilon_{\mu\nu\rho\sigma} (M^{(1)\nu\rho} + M^{(2)\nu\rho})(P^{(1)\sigma} + P^{(2)\sigma})$$

$$\neq W_\mu^{(1)} + W_\mu^{(2)}.$$

Since the two-particle tensor product state is labeled by $2 \times 6 = 12$ quantum numbers, we must find six additional internal quantum numbers in order to fully label our new states. These new labels must not change under arbitrary external inhomogeneous Lorentz transformations and, hence, must be related to the eigenvalues of invariants under these transformations. Four additional invariants are immediately suggested. They are the single-particle invariants $C_1^{(1)}$, $C_2^{(1)}$, $C_1^{(2)}$, and $C_2^{(2)}$. For the remaining two, we take $W_\mu^{(1)}P^{(2)\mu}$ and $W_\mu^{(2)}P^{(1)\mu}$. One can easily verify that the six external invariants, call them $I_1, I_2, \cdots, I_6$, commute among themselves and with the external generators P_μ and $M_{\mu\nu}$. The state of a two-particle system can be written as $| m\, s\, \mathbf{p}\, \lambda; \alpha >$, where $\alpha = (\alpha_1, \cdots, \alpha_6)$ is the eigenvalue of $I = (I_1, \cdots, I_6)$. To see the physical significance of I_5, let us consider the center-of-mass frame of the two particles in which $\mathbf{P}^{(1)} = -\mathbf{P}^{(2)} = \mathbf{p}$. (The numerical values of invariants are, of course, independent of the particular frame used to evaluate them.) Then,

$$I_5 = W_0^{(1)}\,(P_0^{(1)} + P_0^{(2)})$$
$$= p_0\,|\,\mathbf{p}\,|\,\lambda^{(1)};$$

i.e., I_5 is the helicity of the first particle times the product of the total center-of-mass energy and momentum. Similarly,

$$I_6 = p_0\,|\,\mathbf{p}\,|\,\lambda^{(2)}.$$

For fixed $|\,\mathbf{p}\,|$ there exists a correspondence between the two sets of labels $\{I_5, I_6\}$ and $\{\lambda^{(1)}, \lambda^{(2)}\}$; we may use the latter set providing we bear in mind that, under the transformations of $P_+^\uparrow$, the helicities $\lambda^{(1)}$ and $\lambda^{(2)}$ mix in such a way that I_5 and I_6 remain numerically fixed.

It should now be fairly clear how to treat the case of three particles. Namely, we consider the first two particles as a composite system and couple the third particle to it by the procedure outlined above. As a result, we obtain states of the form

$$| m\, s\, \mathbf{p}\, \lambda; \beta >,$$

where β stands for the eigenvalue vector of twelve mutually commuting internal invariants which we take to be:

$$I_1 = P^{(1)} \cdot P^{(1)}, \qquad I_2 = P^{(2)} \cdot P^{(2)},$$
$$I_3 = W^{(1)} \cdot W^{(1)}, \qquad I_4 = W^{(2)} \cdot W^{(2)},$$
$$I_5 = W^{(1)} \cdot P^{(2)}, \qquad I_6 = W^{(2)} \cdot P^{(1)},$$
$$I_7 = P^{(12)} \cdot P^{(12)}, \qquad I_8 = P^{(3)} \cdot P^{(3)},$$
$$I_9 = W^{(12)} \cdot W^{(12)}, \qquad I_{10} = W^{(3)} \cdot W^{(3)},$$
$$I_{11} = W^{(12)} \cdot P^{(3)}, \qquad I_{12} = W^{(3)} \cdot P^{(12)}.$$

We have written

$$a \cdot b = a_\mu b^\mu$$

and

$$P_\mu^{(12)} = P_\mu^{(1)} + P_\mu^{(2)},$$
$$M_{\mu\nu}^{(12)} = M_{\mu\nu}^{(1)} + M_{\mu\nu}^{(2)},$$
$$W_\mu^{(12)} = \tfrac{1}{2}\,\epsilon_{\mu\nu\rho\sigma} M^{(12)\nu\rho}\,P^{(12)\sigma}.$$

The external generators are, of course, just sums of the generators of the three particles.

The general case of an n-particle system is treated similarly. We shall consider the relationship of the external–internal states to the usual tensor product states in a paper presently being prepared.

F. Finite Dimensional Matrix Representations of Infinite Groups

M. M. Saffren

Lately, in our study of relativistic wave equations, we have examined the relationship between the way the complex conjugate of a wave function transforms under Lorentz transformation and the way the wave function itself transforms. While this relationship between the two representations is well-known for the solutions of particular wave equations, the usual treatments of the relationship make it appear that the connection is something characteristic of the wave equation itself, and that, to make the connection, special properties of the wave equation are needed. As we shall show, the relationship follows from some general theorems that are independent of the wave equation and are even independent of its Lorentz invariance. The language and notation of group theory will be used without definitions of terms; this information is available in Ref. 16.

For finite groups, the connection between the transformation of a basis and that of the complex conjugate basis has long been known to mathematicians and physicists alike. The matter is made simple by virtue of the equivalence of all representations of finite groups to

unitary representations. For infinite groups, however, not all finite dimensional representations are unitary. Nevertheless, as we show here, all the theorems that hold for finite groups and have to do with complex conjugate representations hold for finite-dimensional representations of infinite groups as well. From the theorems we prove, which hold for any group, the theorems regarding the Lorentz group follow as an immediate consequence.

For finite groups one can show that, if the complex conjugate of a representation is equivalent to the representation itself, then either the representation is equivalent to a real representation (the basis can be chosen to consist of real functions), or the representation is "pseudo-real" (the basis cannot be chosen to consist of real functions alone, but, if a function appears in the basis, so does its complex conjugate which is distinct from it). Of course, it may be that the representation is "complex" (the complex conjugate basis transforms according to a representation that is entirely inequivalent to the original).

The same three alternatives (real, pseudo-real, and complex) hold for the finite-dimensional irreducible representations of infinite groups. But, in what follows, we have no further need to allude to the basis and so we consider only the matrices of the representation. If $M(R_i)$ denotes the matrix representative of the element R_i of the abstract group being represented, then, as is well-known (and is nearly obvious), the following three matrices are also representatives: $M^{-T}(R_i)$, the inverse transposed matrix; $M^+(R_i)$, the complex conjugate matrix; and $M^{-T+}(R_i)$, the complex conjugate of the inverse transposed matrix. Because for finite groups the representation matrices may be chosen to be unitary, there are only two matrices to consider: M and M^+; the other two are identical to these. For infinite groups this degeneration does not occur; thus, besides asking about the relationship between M and M^+, we can also ask about the relationship between M and the other two matrices M^{-T} and M^{-T+}. The resulting relationships are expressed in the three theorems discussed below. Aside from any intrinsic interest these theorems may have, they will prove useful in providing a simpler treatment than has heretofore been obtained of the behavior of representations of the Lorentz group under the discrete operations of charge conjugation, parity, and time inversion.

Theorem I. Suppose that M and M^{-T+} are equivalent representations of a group and that M is irreducible. Then, M is also equivalent to a representation N and $NDN^{T+} = D$, where D is diagonal, $D^2 = I$, and I is the

unit matrix. For finite groups $D = I$. Proof of this theorem follows:

Proof. If a representation M is equivalent to its inverse adjoint M^{-T+}, then a matrix S exists that satisfies

$$M^{-T+} = SMS^{-1}. \qquad (1)$$

Clearly, $S^{-T+}S$ commutes with M, and, if M is irreducible, Schur's lemma requires that $S^{+T} = \lambda S$, where λ is a number. Clearly, λ must have unit modulus. Redefining $\lambda^{1/2}S$ as S, the redefined S is seen to be Hermitian. Now S has, as does any matrix, a unique polar form UH, where U is unitary and H is Hermitian and positive. Thus, $S = S^{+T}$ requires that $UH = HU^{+T}$, but, since the polar form is unique, this requires that

$$U^2 = I, \qquad (2a)$$

$$H = UHU^{+T}. \qquad (2b)$$

Now let $S' = H^{-1/2}UH^{1/2}$ and $M' = H^{1/2}MH^{-1/2}$. Eq. (1) is now $M'^{-T+} = S'M'S'^{-1}$. The matrix S' satisfies both $S'S' = I$ and $S'S'^{-T+} = I$, as is seen by using Eq. (2). We conclude S' is both unitary and Hermitian.

We now define S'' to be the diagonalized form of S' (i.e., $S' = VS''V^{+T}$) and M'' as $M'' = V^{+T}M'V$. Eq. (1) can now be written as

$$M''^{-T+} = S''M''S'', \qquad (3)$$

where S'' is a diagonal matrix whose non-zero elements are either $+1$ or -1, so that $S''^2 = I$. With $S'' = D$ and $M'' = N$, the theorem is proved.

Consequences. The following are consequences of Theorem I:

(1) Obviously, if $S'' = \pm I$, all the M'' are unitary. If $S'' \neq \pm I$, those matrices that commute with S'' are unitary.

(2) If $S'' \neq \pm I$, not all the M can be unitary. If they were, S'' would commute with them all and the representation could not be irreducible.

(3) If M is not irreducible, the condition of equivalence of M with its inverse adjoint does not require M to be decomposable. This is shown in the following example: Let

$$T_n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix},$$

where n is an integer. Then,

$$T_nT_m = T_{m+n}.$$

These matrices are reducible. Moreover,

$$T_n^{-T+} = \begin{bmatrix} 1 & 0 \\ -n & 1 \end{bmatrix},$$

$$T_n^{-T} = S T_n S^{-1},$$

where

$$S = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}.$$

(4) If S is Hermitian but not positive, then M is not required to be decomposable if it is reducible. This is shown by the example given above.

(5) If S is Hermitian and positive, reducibility implies decomposability. This always happens for finite groups.

(6) If M is real and $s^2 = S$ defines s, then

$$s M'^T s^+ s^+ M' s = s;$$

thus, $Q = s^+ M' s$ is orthogonal though not necessarily real.

(7) If M' is orthogonal though not real,

$$M'^+ = S'' M' S'';$$

then, $Q = s^+ M' s$ is real and orthogonal.

(8) If M' is symmetric though not real, then

$$M'^+ S' M' = S',$$

$$s^+ M'^+ s s M' s^+ = I.$$

If $Q = s M' s^+$, $Q^{+T} Q = I$ and Q is unitary.

Theorem II. Under the hypothesis of Theorem I, with M^{-T+} replaced by M^+, M is equivalent to either a real or a pseudo-real representation. The proof for this follows:

Proof. Suppose

$$M^+ = SMS^{-1}. \qquad (4)$$

Then, for S in polar form UH, and with $M' = H^{\frac{1}{2}} M H^{\frac{1}{2}}$ and $S' = H^{+\frac{1}{2}} U H^{\frac{1}{2}}$, Eq. (4) becomes

$$M'^+ = S' M' S'^{-1}.$$

If M' is irreducible, $S'^T S'$ which commutes with M' is proportional to the unit matrix. The constant of proportionality is real, and we may define S'' in such a way that

$$S'' = \pm S''^{-1+}. \qquad (5)$$

The sign in Eq. (5) is the sign of the constant.

For S'' in polar form, Eq. (5) becomes

$$H^+ U U^+ = \pm U^T H^{-1} U^+.$$

The uniqueness of the polar form now shows that $UU^+ = \pm I$ and $H^+ = U^T H^{-1} U^+$. By using these relations we find

$$S''^{+T} S'' = I.$$

Thus, S'' is unitary. This, with Eq. (5), shows that $S'' = \pm S''^T$. When $S'' = S''^T$, the representation M' is equivalent to a real representation (see Appendix A). When $S'' = - S''^T$, M' is not equivalent to a real representation, but instead to a pseudo-real representation (see Appendix B).

Theorem III. Under the hypothesis of Theorem I, with M^{-T+} replaced by M^{-T}, M is equivalent to either an orthogonal or a symplectic representation. (Theorem III has no analogue for representations of finite groups.) The proof of this theorem is as follows:

Proof. Suppose

$$M^{-T} = S M S^{-1}.$$

Clearly, $S^{-T} S$ commutes with M. If M is irreducible, $S^{-T} S$ is proportional to the unit matrix, and this requires $S = \pm S^T$. The polar form UH of S shows $U = \pm U^T$ and $H = U^+ H^T U'^T$. With $S' = H^{-\frac{1}{2}} U H^{\frac{1}{2}}$ and $M' = H^{\frac{1}{2}} M H^{-\frac{1}{2}}$, we have $M'^{-T} = S' M' S'^{-1}$, and S' can be shown to be unitary. Now S' is either symmetric or antisymmetric. When S' is symmetric, M' is equivalent to an orthogonal transformation (see Appendix A). When S' is antisymmetric, M' is equivalent to a symplectic transformation (see Appendix B).

Appendixes to Theorems II and III

Appendix A. Here we shall prove the well-known theorem that, if S is a unitary symmetric matrix, then $S = VV^T$, where V is unitary. This result must be used to complete Theorems II and III. In Theorem II, $M'^+ = S' M' S'^{-1}$ and $S'^T = \pm S'$. If $S'^T = \pm S'$, then, as we claim, $S' = VV^T$ and so $V^+ M'^+ V = V^T M' V^+$ and $V^T M' V^+$ is real. In Theorem III, $M'^{-T} = S' M' S'^{-1}$ and $S'^T = \pm S'$. If $S'^T = S'$, then $V^{-1} M'^{-T} V = V^T M' V^{-T}$ and $V^T M' V^{-T}$ is orthogonal.

We now prove $S = VW^T$. Because S is unitary, it can be diagonalized by a unitary matrix W; i.e., $W^{+T} S W = S_d$, where S_d is diagonal. Consider now the symmetric matrix $W^{+T} S W^+$, which is equal to $S_d W^{+T} W^+$. Because this matrix

is symmetric, S_d commutes with $W^{+T}W^+$. If $(S_d)_i$ is a diagonal element of S_d and if $(S_d)_j$ is another diagonal element equal to it, then define $(S_d)^{\frac{1}{2}}$ so that $(S_d)_i^{\frac{1}{2}} = (S_d)_j^{\frac{1}{2}}$. With $(S_d)^{\frac{1}{2}}$ defined this way, $(S_d)^{\frac{1}{2}}$ commutes with $W^{+T}W^+$. Thus,

$$S = WS_d^{\frac{1}{2}} W^{+T}W^+S_d^{\frac{1}{2}} W^T.$$

If we define V as $WS_d^{\frac{1}{2}} W^{+T}$, then

$$S = VV^T.$$

Appendix B. We shall now establish one of many possible canonical forms for unitary skew-symmetric matrices. We show that, if S is such a matrix, $S = VKV^T$, where

$$K = \begin{bmatrix} 0 & -iI \\ iI & U \end{bmatrix}$$

and V is unitary. Then, a pseudo-real representation has the canonical form

$$M'''^+ = KM''K,$$

and a symplectic representation has the form

$$M'''^T (iK)M'' = (iK).$$

Since S is unitary, it can be diagonalized by a unitary matrix W. Therefore, $W^{+T}SW = S_d$. Now

$$W^{+T}SW^+ = S_dW^{+T}W^+$$

is skew, so that S_d anticommutes with $W^{+T}W^+$. A permutation matrix $P(P^2 = I)$ rearranges S_d so that it has the form

$$S_d' = \begin{bmatrix} (\lambda)_m & 0 \\ 0 & (\lambda)_n \end{bmatrix},$$

where $(\lambda)_m$ is an $m \times n$ diagonal matrix, each of whose diagonal elements has a phase ϕ, $0 \leq \phi < \pi$, and $(\lambda)_n$ is an $n \times m$ diagonal matrix, each of whose diagonal elements has a phase ϕ, $\pi \leq \phi < 2\pi$. The matrix $A = PW^{+T} WP$ is unitary and anticommutes with S_d'. If we write this in partitioned form,

$$\begin{bmatrix} (\lambda)_mA_n & (\lambda)_mA_{12} \\ (\lambda)_nA_{21} & (\lambda)_n A_{22} \end{bmatrix} = \begin{bmatrix} -A_{11}(\lambda)_m & -A_{12}(\lambda)_n \\ -A_{12}(\lambda)_m & -A_{22}(\lambda)_n \end{bmatrix}.$$

We see that, since $(\lambda_m)_i (A_{11})_{ij} = -(A_{11})_{ij} (\lambda_m)_j$, $A_{11} = 0$. Similarly, $A_{22} = 0$. Thus, A, which is symmetric, has the form

$$\begin{bmatrix} 0 & A_{12} \\ (A_{12})^T & 0 \end{bmatrix}.$$

The matrix A_{12} is an $m \times n$ matrix, and we will show that $m = n$. Now, because A is unitary, $(A_{12}) (A_{12})^{T+} = (I)_m$ and $(A_{12}^T) (A_{12})^+ = (i)_n$. Consider the first of these two equations. It represents $2m^2 - m$ real equations for the $2mn$ quantities $(A_{12})_{ij}$ and $(A_{12})_{ij}^+$, so $2m^2 - m \leq 2mn$. Similarly, the second equation requires $2n^2 - n \leq 2mn$. Thus, $m = n$, and we see that

$$W^{+T}SW^+ = \begin{bmatrix} -\lambda & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} 0 & A_{12} \\ (A_{12})^T & 0 \end{bmatrix}$$

$$= \begin{bmatrix} (i\lambda)^{\frac{1}{2}} & 0 \\ 0 & (i\lambda)^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & A_{12}^T \end{bmatrix} \begin{bmatrix} 0 & -iI \\ iI & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} I & 0 \\ 0 & A_{12} \end{bmatrix} \begin{bmatrix} (i\lambda)^{\frac{1}{2}} & 0 \\ 0 & (i\lambda)^{\frac{1}{2}} \end{bmatrix}.$$

Therefore, if W' is defined as

$$W' = \begin{bmatrix} -(i\lambda)^{\frac{1}{2}} & 0 \\ 0 & (i\lambda)^{\frac{1}{2}} (A_{12})^T \end{bmatrix},$$

then W' is unitary and

$$S = W' \begin{bmatrix} 0 & -iI \\ iI & 0 \end{bmatrix} W'^T.$$

References

1. Stumpf, H. J., "Thermal Radiation Loss From Vortex Tubes," *RS 36-14,* pp. 115–117, Jet Propulsion Laboratory, Pasadena, California, May 1, 1962.

2. Meghreblian, R. V., *Thermal Radiation in Gaseous Propulsion Reactors,* Technical Report No. 32-139, Jet Propulsion Laboratory, Pasadena, California, July 24, 1961.

3. Stumpf, H. J., *Fission Fragment Energy Loss From Vortex Tubes,* Technical Report No. 32-188, Jet Propulsion Laboratory, Pasadena, California, March 12, 1962.

References (Cont'd)

4. Fastie, W. G., "Instrumentation for Ultraviolet Rocket Spectrophotometry," *Journal of Quantitative Spectroscopy and Radiative Transfer*, Vol. 3, pp. 507–518, 1963.

5. Chamberlain, J. W., *Physics of the Aurora and Airglow*, Chapter 11, Academic Press, New York, 1961.

6. Malitson, H. H., Purcell, J. D., Tousey, R., and Moore, C. E., "The Solar Spectrum from 2635 to 2085 A," *Astrophysical Journal*, Vol. 132, pp. 746–766, 1960.

7. Tousey, R., "The Extreme Ultraviolet Spectrum of the Sun," *Space Science Reviews*, Vol. 2, pp. 3–69, 1963.

8. Nicholls, R. W., "Transition Probabilities for Aeronomically Important Spectra," *Annales de Geophysique*, Vol. 20, 1964.

9. Jursa, A. S., Tanaka, Y., and LeBlanc, F., "Nitric Oxide and Molecular Oxygen in the Earth's Upper Atmosphere," *Planetary and Space Science*, Vol. 1, pp. 161–172, 1959.

10. Molmud, P., "The Expansion of a Rarefied Gas into a Vacuum," *Physics of Fluids*, Vol. 3, p. 363, 1960.

11. Keller, J. B., "On the Solution of the Boltzmann Equation for Rarefied Gases," *Communications on Pure and Applied Mathematics*, Vol. 1, p. 275, 1948.

12. Dawson, J., "Nonlinear Electron Oscillations in a Cold Plasma," *Physical Review*, Vol. 113, p. 383, 1959.

13. von Roos, O., *Plasma Theory of Electron–Phonon Interaction: II*, Technical Report No. 32-106, Jet Propulsion Laboratory, Pasadena, California, June 25, 1961.

14. Lindhard, J., *Det Kongelige Danske Videnskabernes Selskab, Matematisk-Fysiske Meddeleser*, Vol. 28, p. 8, 1954.

15. Pines, D., *Elementary Excitations in Solids*, W. A. Benjamin, Inc., New York, 1963.

16. Hamermesh, *Group Theory*, Addison–Wesley Publishing Company, Inc., Cambridge, Massachusetts, 1963.

TELECOMMUNICATIONS DIVISION

XVII. Communications Elements Research

A. Low-Noise Amplifiers

C. T. Stelzried, W. V. T. Rusch, and R. Brantner

1. Millimeter-Wave Circuit Element Evaluation,
C. T. Stelzried and W. V. T. Rusch

Millimeter-wave components and techniques are being investigated to ascertain the future applicability of this frequency range to space communications and tracking. This involves the development of instrumentation for accurate determinations of insertion loss, VSWR, power and equivalent noise temperature of passive elements, and gain and bandwidth of active elements at millimeter wavelengths.

To bring together state-of-the-art millimeter-wave circuit elements and to evaluate their use in a system, we have built a radiotelescope consisting of a 60-in. antenna and a superheterodyne radiometer. The radiotelescope was used to observe the 90-Gc temperature of the Moon

during the December 30, 1963 eclipse. This experiment was a joint effort by personnel from JPL and the Electrical Engineering Department of the University of Southern California. As part of the radio telescope alignment and testing, eleven drift curves of the Sun were recorded on December 29, 1963, and the Sun temperature was determined as a function of zenith angle. Atmospheric loss at the zenith on this particular day was found to be about 0.50 db at 90 Gc. In this summary, the radiometer waveguide calibrations and performance capability are presented.

a. 90-Gc radiotelescope observations. A 90-Gc radiotelescope (Ref. 1) was assembled and operated at the JPL Mesa Antenna Range to measure the temperature of the Moon during the lunar eclipse on December 30, 1963. During the previous afternoon, eleven drift curves of the Sun were made to measure atmospheric absorption and atmospheric self-emission and to align the optical and radio axes of the antenna. Thin brown paper was

used in front of the antenna reflector to prevent heat damage. A typical drift curve of the Sun is shown in Fig. 1. For this curve, the zenith angle was 75.0 deg, the local time was 15:20 PST, and the radiometer output time constant was 23 sec. The approximate 1400°K excess noise injected from the gas tube was supplied through a 10-db coupler. The ratio between the gas tube and the deflection due to the Sun was determined and plotted as a function of the secant of the zenith angle (Fig. 2). Assuming a flat Earth (Fig. 3) and an atmospheric loss L_0 at zenith, the attenuation at any zenith angle z (accurate for angles less than about 80 deg) is

$$L = L_0^{\sec z} \qquad (1)$$

Fig. 2. Measured ratio of 90-Gc Sun temperature to gas tube temperature versus secant z
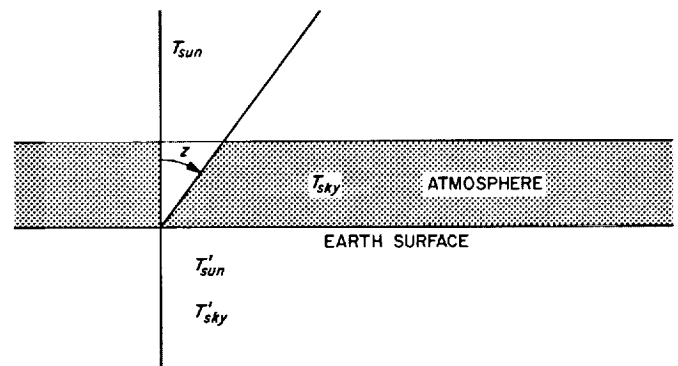
Fig. 3. Geometry used with flat Earth for computation of atmospheric attenuation

The temperature of the Sun as seen by the radiotelescope is related to the equivalent temperature of the sky, T_{sky}, and that seen if there were no attenuation, T_{sun}, by

$$T'_{sun} = \left(1 - \frac{1}{L}\right) T_{sky} + \frac{1}{L} T_{sun} \qquad (2)$$

A drift curve of the Sun with the radiotelescope measures

$$\Delta T'_{sun} = T'_{sun} - \left(1 - \frac{1}{L}\right) T_{sky} \qquad (3)$$

So that

$$T_{sun} = L \Delta T'_{sun} \qquad (4)$$

Fig. 1. Typical 90-Gc radiotelescope Sun drift curve

The atmospheric attenuation at zenith is [Eqs. (1) and (4)],

$$L_0 = \left[\frac{\Delta T'_{sun}(z_1)}{\Delta T'_{sun}(z_2)}\right]^{\frac{1}{\sec z_2 - \sec z_1}}$$

or

$$L_0(\text{db}) = \frac{10}{\sec z_2 - \sec z_1}\left[\log \Delta T'_{sun}(z_1) - \log \Delta T'_{sun}(z_2)\right] \tag{5}$$

The scale factor for T'_{sun} cancels out and from Fig. 2 with secant $z_2 = 5$, secant $z_1 = 0$ (for 12-29-63 at 90 Gc),

$$L_0 (\text{db}) \simeq \frac{10}{5} (\log 3.00 - \log 1.68) = 0.50 \tag{6}$$

The focusing of the antenna subreflector for the Sun measurement and the Moon eclipse measurements was accomplished by maximizing the noise power from the Sun. Although this is not a point source and does not produce a proper focus, no other way was possible at the time because of the absence of a signal source. Figs. 4(a) and 4(b) show the antenna pattern measured after a CW signal source had become available. The antenna will be refocused using the CW signal source, and the new patterns will be published at a future date.

From 6:30 pm December 29, 1963 to 6 am December 30, 1963, forty-three curves of the Moon were recorded in an effort to measure the temperature change during the eclipse. Prior to recording each drift curve, the radiometer was calibrated by switching between ambient and hot calibrated terminations. Fig. 5 shows a typical drift curve (Record 12) with the hot and ambient termination calibrations. Conditions for this recording were

Time constant	= 10 sec
Local time	= 9:21 pm PST
Zenith angle	= 33.25 deg
Hot load temperature	= 327°C
Ambient load temperature	= 22°C
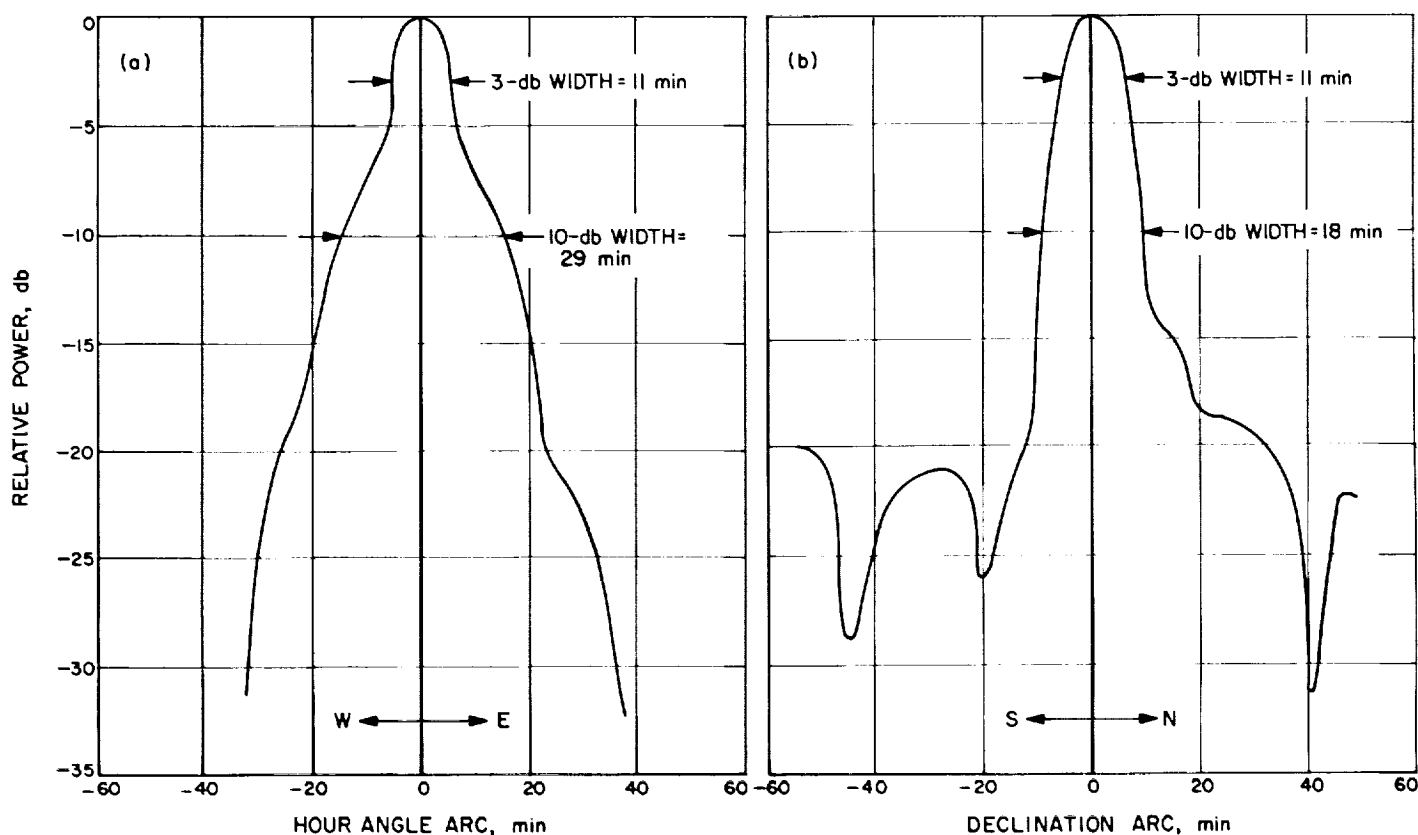
Fig. 4. Antenna patterns in hour angle and declination

A preliminary hand reduction of the lunar temperature measurements showed a standard deviation of the individually measured points about the best least-squares curve fit of 39°K. The average of the six measured lunar temperatures between the time that totality began and ended was 10°K below the curve fit.

b. Digital computer programming for Moon drift curve data reduction. A digital computer program is being set up to facilitate the reduction of the Moon drift curve data for the present and future measurements. The analog data from the drift curves have been digitized and will be compared on a least-squares error basis with a theoretical drift curve. The fit will be adjusted to give outputs in the program of the maximum temperature magnitude and time.

Fig. 6 shows the geometry used in the calculation of a theoretical Moon drift curve using

$$T_A (\theta, \phi) = k \int \int_s G (\theta, \phi) d\Omega$$

or

$$T_A (\theta, \phi) = k \int \int_s G (\theta, \phi) \sin \theta \, d\theta \, d\phi$$

(7)

The antenna patterns were measured in the N-S and E-W directions (Fig. 4) and can be approximated elsewhere by

$$G (\theta, \phi) = \begin{cases} F_E \cos^2 \phi + F_N \sin^2 \phi & ; 0 \le \phi \le \pi/2 \\ F_N \sin^2 \phi + F_W \cos^2 \phi & ; \pi/2 \le \phi \le \pi \\ F_W \cos^2 \phi + F_S \sin^2 \phi & ; \pi \le \phi \le 3\pi/2 \\ F_S \sin^2 \phi + F_E \cos^2 \phi & ; 3\pi/2 \le \phi \le 2\pi \end{cases}$$

(8)

Fig. 5. Typical 90-Gc radiometer Moon drift curve
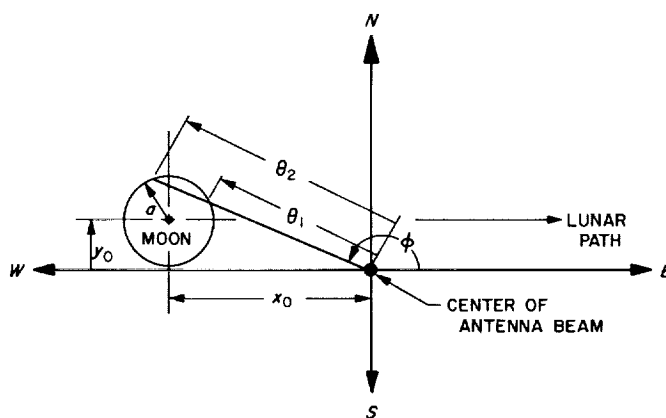
Fig. 6. Geometry of Moon passage over antenna beam used in theoretical Moon drift curve calculation

where F_E, F_W = relative gain in the East and West directions, respectively (see Fig. 4a), and F_N, F_S = relative gain in the North and South directions, respectively (see Fig. 4b). A theoretical drift curve can be obtained by integrating over the surface of the Moon with a given value for y_0 and moving x_0 over the range of interest. If ϕ_0 is the value of ϕ to the Moon center,

$$T_A \simeq \left\{ \begin{array}{l} \displaystyle\int_{\phi_1}^{\phi_2} d\phi \int_{\theta_1(\phi)}^{\theta_2(\phi)} G(\theta, \phi) \sin\theta \, d\theta \, d\phi; \ (x_0^2 + y_0^2)^{1/2} \geq a \\[2em] \displaystyle\int_0^{2\pi} d\phi \int_0^{\theta_2(\phi)} G(\theta, \phi) \sin\theta \, d\theta \, d\phi; \ (x_0^2 + y_0^2)^{1/2} \leq a \end{array} \right\}$$

$$(9)$$

where

$$\phi_1 = \phi_0 - \sin^{-1}\left[\frac{a}{(x_0^2 + y_0^2)^{1/2}}\right]$$

$$\phi_2 = \phi_0 + \sin^{-1}\left[\frac{a}{(x_0^2 + y_0^2)^{1/2}}\right]$$

This has been computed for a = 16.50 min arc (the radius of the Moon on December 30, 1963) and $y_0 = 0$ on the digital computer and is shown in Fig. 7. Additional theoretical curves will be computed using various values of y_0 to indicate the sensitivity to antenna pointing. This technique has been used by others but with an analytic expression for $G(\theta, \phi)$ when θ is small (Ref. 2).
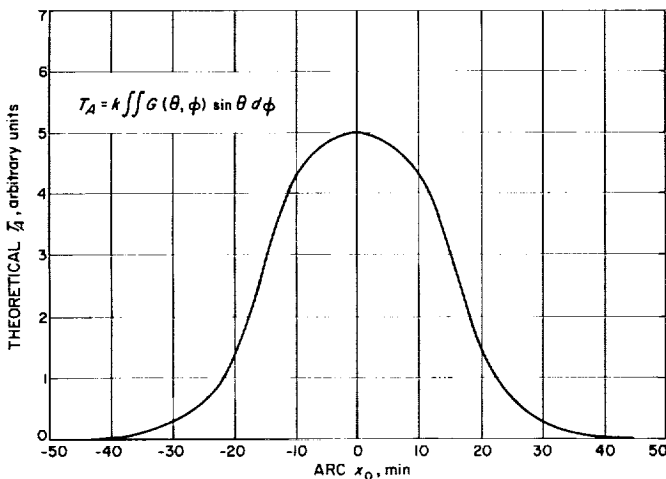
Fig. 7. Theoretical Moon drift curve for 90-Gc 60-in. searchlight radio telescope

The actual Moon drift curves have not yet been fitted to the theoretical curves.

c. Radiometer calibrations. A diagram of the waveguide plumbing for the 90-Gc radiometer as used during the December 30, 1963 Moon eclipse is shown in Fig. 8. The measured insertion losses of the critical components are shown on the figure. The equivalent input noise temperature of the hot load has been calibrated and is tabulated later (see Sect. A3 of this summary). To a good approximation

$$(T_H' - T_0) \simeq 0.87 (T_H - T_0) \qquad (10)$$

where T_0 is the ambient temperature, T_H the actual temperature of hot load (measured), and T_H' the hot load temperature calibrated at output of oven.

Then with $L_2 \simeq 1.43$ db insertion loss (Fig. 8) at ambient temperature,

$$(T_H'' - T_0) \simeq 0.62 (T_H - T_0) \qquad (11)$$

where T_H'' is the hot load temperature calibrated at output of waveguide switch. Drift curve temperature measurements $\Delta T_A''$ can now be calibrated at the output of the waveguide switch. To refer the measurements to the horn input (Fig. 8),

$$\Delta T_A' = L_3 \Delta T_A'' \simeq 1.10 \, \Delta T_A'' \qquad (12)$$

The performance of the radiometer on December 29, 1963 can be estimated from the calibrations in Fig. 5, where 2 divisions represent $(327-22)°K$, actual measured temperatures. From Eq. (12), referred to the output of the waveguide switch,

$$\text{Calibration} \simeq \frac{0.62 \, (327-22)}{2} \simeq 100 \left(\frac{°K}{\text{division}}\right)$$

The peak-to-peak jitter is approximately $100 \times 0.3 \simeq 30°K$. The theoretical jitter is about $3°K$ rms, based on a 10-sec time constant, approximately $20,000°K$ receiver temperature and 10-Mc bandwidth. The noise jitter in the radiometer output thus appears reasonable, but the drift or change in base line is excessive and degrades the radiometer performance seriously (Fig. 5). It has been found that this drift was caused by variations of a large in-phase signal at the reference frequency in the Dicke
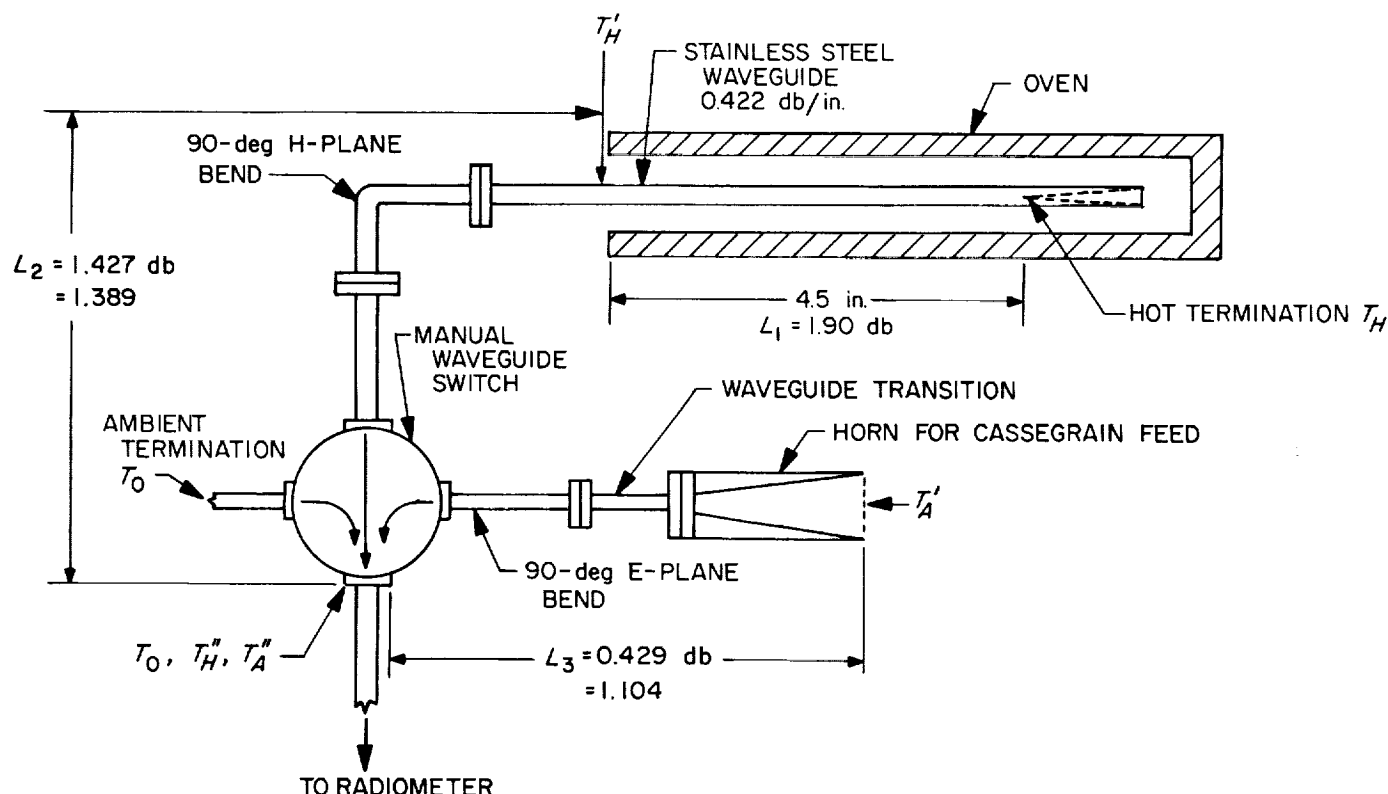
Fig. 8. 90-Gc millimeter radiometer temperature calibrations

system caused by change in VSWR of the ferrite switch affecting the receiver input mixer performance. Adding an isolator after the ferrite switch nearly eliminated this effect with a substantial increase in stability. To facilitate measuring long-term and short-term gain stability of the radiometer, an automatic sequencing system has been implemented for the noise source and the waveguide switch. The present radiometer long-term stability is shown by the recording shown in Fig. 9. The time constant is 1.5 sec with a 20,000°K receiver temperature and 10-Mc bandwidth. The waveguide input switch is switched automatically between ambient and hot temperature terminations, and the noise source is (approx 300°K) periodically turned on. Fig. 10 shows recordings of the short-term gain stability with 1.5 and 12 sec time constants for a portion of a 12 on-off noise source sequence. These were taken at a faster recorder speed with the noise source (approx 300°K) manually turned off and then on for equal periods of time. The digital recorder output from an integrating voltmeter is shown plotted in Fig. 11 for the same period of time. The digital voltmeter input, supplied from the detector after the 1.5-sec time constant (as used for the output shown in Fig. 10), integrates and prints out every 10 sec. The data from the digital recorder have been further analyzed as follows:

The averages of the data points for all the odd and even runs, nominally 294°K and 594°K, respectively, are used to establish the scale. Table 1 shows for each run the run identification number (1 to 12), the number of digital data points (8 to 13), the average temperature

Table 1. Summary of 90-Gc radiometer digital output with 300°K noise source periodically switched on (even numbers) and off (odd numbers)

Run No.	Number of digital data points	Average measured temperature, °K	Deviation rms, °K	Measurement error, °K
1	11	294.48	1.55	−0.48
2	8	591.24	1.35	2.76
3	10	288.11	1.06	5.89
4	12	591.32	2.61	2.86
5	13	294.21	1.66	−0.21
6	11	596.17	1.65	−2.17
7	12	293.76	1.26	0.24
8	11	592.35	2.77	1.65
9	13	298.72	0.42	−4.72
10	8	595.13	1.34	−1.13
11	8	292.95	0.35	1.05
12	10	597.93	2.52	−3.93
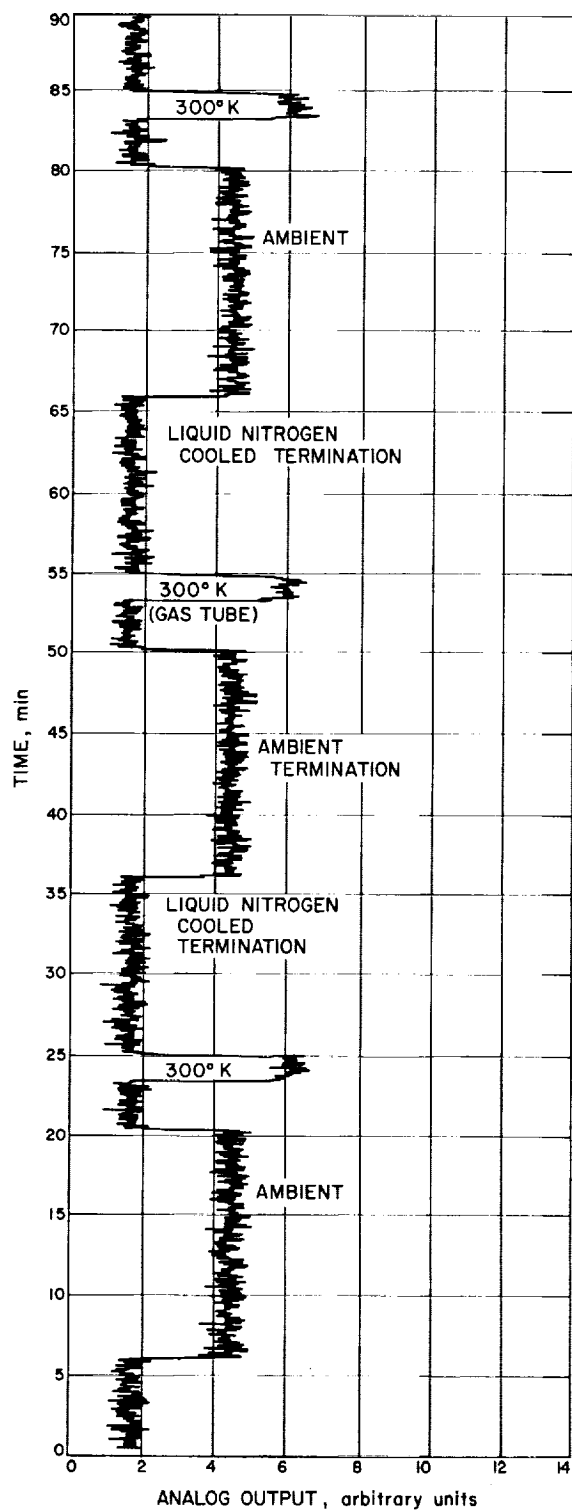
Fig. 9. 90-Gc radiometer long-term stability recording with automatic waveguide and noise source switching
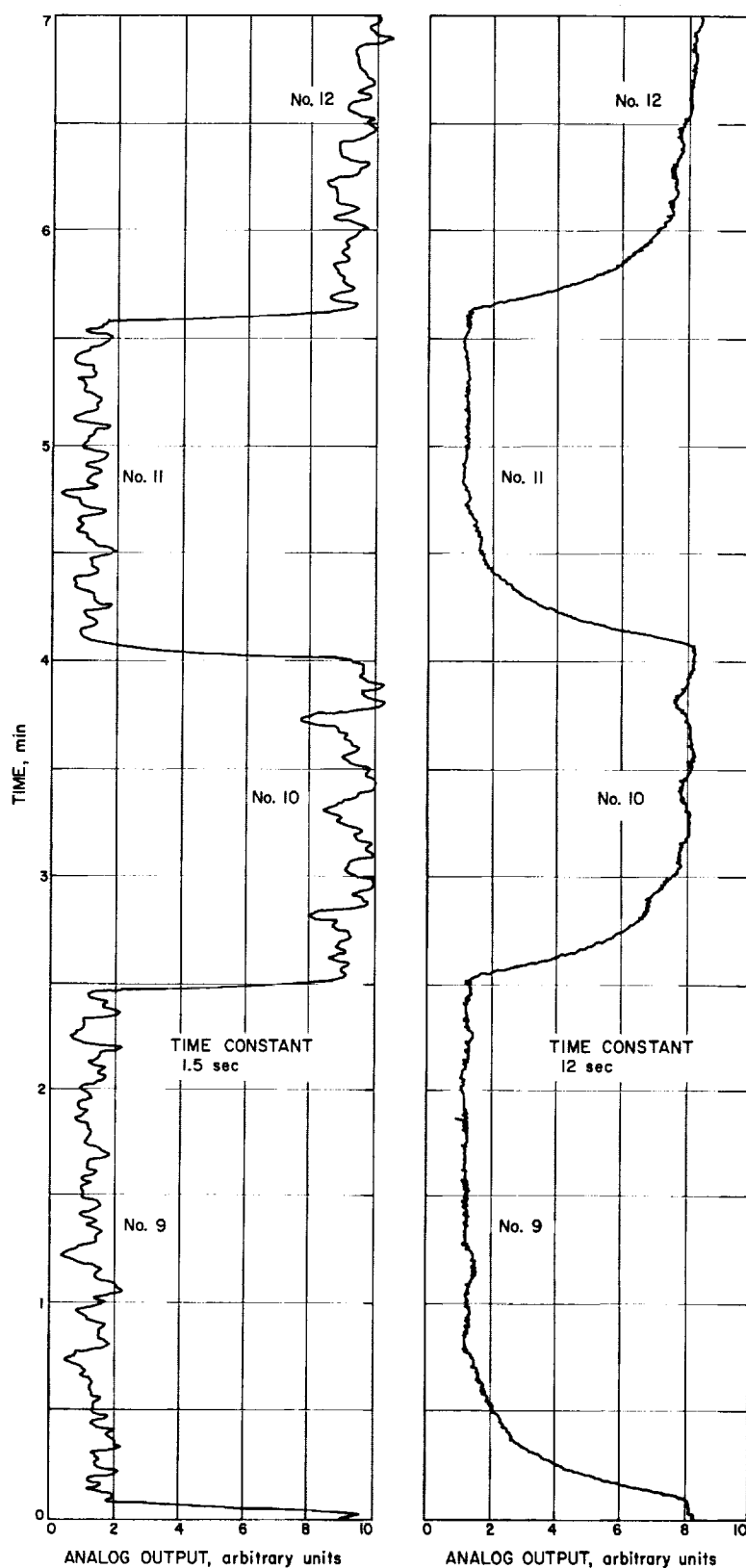
Fig. 10. 90-Gc radiometer short-term stability with 1.5- and 12-sec time constants
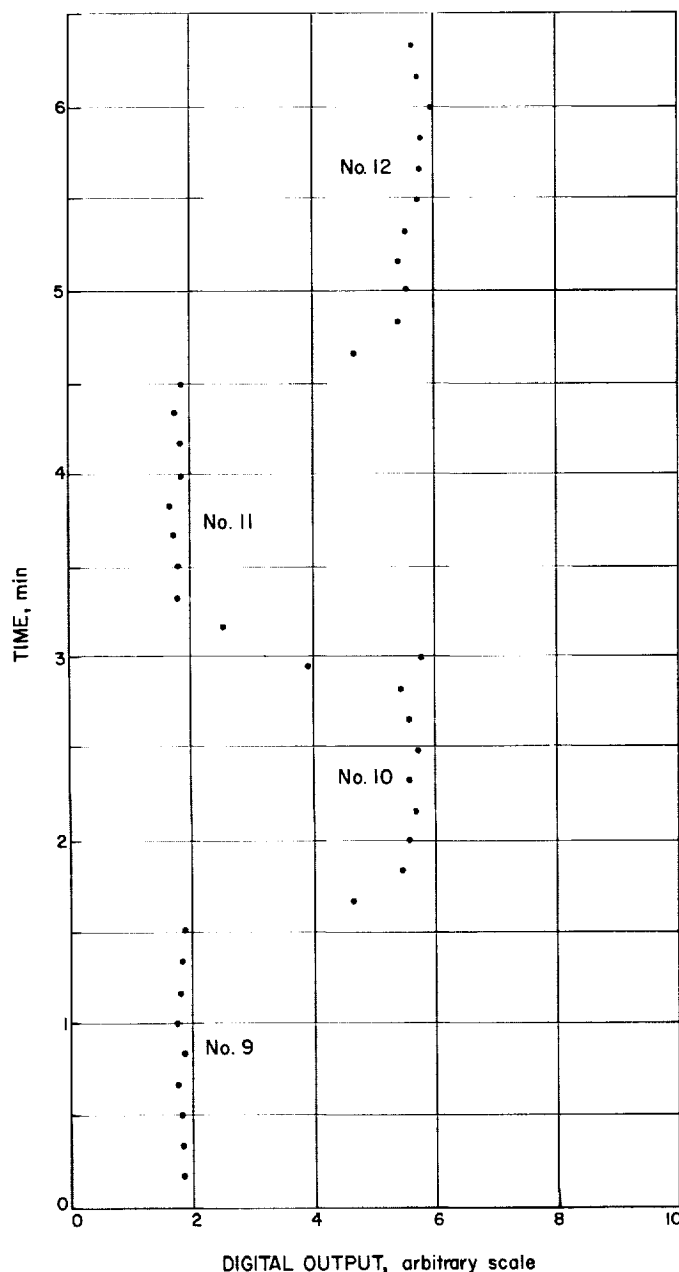
Fig. 11. 90-Gc radiometer short-term stability digital printout with 10-sec integrating time

Fig. 12. Graph of digital measurement error versus run number

absolute values of the measurement errors, 2.25°K. This value would decrease with an increase in data points for each run (in this case, the average number of data points is 10.7 for each run) unless there was a systematic gain change (non-random noise).

2. Solid State Circuits, R. Brantner

Intermediate frequency amplifiers using field effect transistors are being investigated for possible application as low-noise amplifiers where space and power requirements are severe. An experimental 10-Mc amplifier has been built and tested.

A major source of noise in bipolar transistors is shot noise associated with the flow of carriers across the emitter-base and collector-base junctions. No such junctions are present in the flow path of carriers in field-effect transistors (FET), so that a lower noise output may be expected. For this reason, an attempt is being made to use FET's in a low-noise IF amplifier operating at 10 Mc. The first model constructed contained five stages of amplification plus an impedance conversion stage, all of which were fairly conventional. Gains of up to 100 db were obtained, but with a bandwidth far too narrow to be useful. After loading the amplifier stages for greater bandwidth, a noise figure of 4.3 db was obtained at a gain of about 60 db and a bandwidth of about 1 Mc. Tests on the preliminary IF amplifier indicated that a much more carefully shielded chassis is required, and that neutralization should be used at least on the input stage and probably on all amplifier stages. Accordingly, an improved model is now under construction incorporating these features.

(288°K to 299°K for the odds, and 591°K to 598°K for the evens), the rms deviation (0.35°K to 2.77°K), and the measurement error (−0.21°K to 5.89°K). The measurement error is defined as the difference between the nominal temperature and the average measured temperature for each run. The measurement error for each run is plotted in Fig. 12. One number to characterize the performance of the radiometer might be the average of the

The FET's used to date are type TA 2330, produced by RCA. They are of the insulated gate variety in which a thin film of silica electrically isolates the metal film gate from the channel. This type of FET was chosen because of the lack of voltage-dependent input capacity as compared to junction-type FET's, and because of its higher transconductance. Even so, it is doubtful whether an FET IF amplifier can perform as well as a good low-noise vacuum tube (VT) IF amplifier, since the transconductance of the FET's is about $\frac{1}{25}$ that of the vacuum tubes used in our better VT IF amplifiers.

Work is continuing on the FET IF amplifier in optimizing its low-noise gain characteristics. It is expected that, using FET's, an IF amplifier may be produced that has better noise-gain characteristics than present bipolar transistor IF amplifiers and requires much less power and space than does the average VT IF amplifier.

3. Temperature Calibration of Microwave Terminations, C. T. Stelzried

Microwave terminations of known calibrated temperatures are needed for radiometry, antenna temperature measurements, and amplifier performance evaluation (Ref. 3). The available thermal noise power from the termination is given by kTB (assuming $hf/kT >> 1$) where k is the Boltzmann's constant (1.38×10^{-23} joules/°K), T is the temperature (°K), and B the bandwidth (cps).

In this summary, the equation governing the equivalent input noise temperature for a transmission line with a termination of specified temperature is solved and the solutions tabulated for various combinations of transmission line temperature distributions. A computer program has been developed for arbitrary transmission line temperature distributions and used to evaluate a 90-Gc hot termination.

The transmission line input noise temperature is derived as a function of frequency, accounting for the variation in losses with frequency. Appropriate measurements have been made for a commercial coaxial transmission line and the input noise temperature computed and graphed as a function of frequency.

a. Transmission line with arbitrary temperature distribution. The equivalent input noise temperature T' for a transmission line of length l and insertion loss L, with a termination at a temperature T, temperature distribution $T_L(x)$ and losses independent of temperature, is (Ref. 4)

$$T' = \frac{T + \frac{\mathcal{L}}{l} \int_0^l \exp\left(\frac{\mathcal{L}x}{l}\right) T_L(x)\, dx}{L} \tag{1}$$

where

$$\mathcal{L} = \frac{L\,(\mathrm{db})}{10 \log_{10} e} \simeq 0.2303\, L\,(\mathrm{db})$$

x = distance from termination

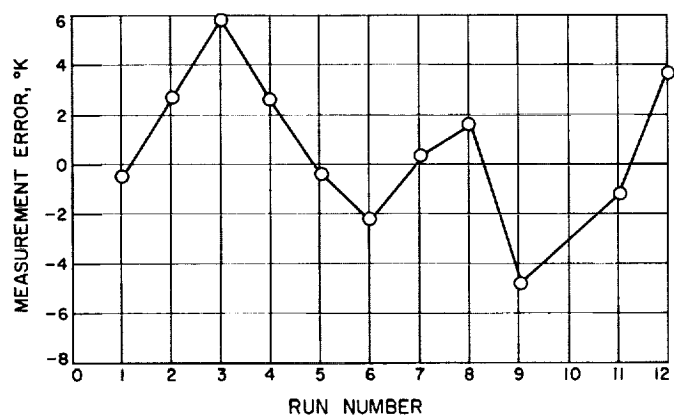This equation can be integrated directly if the temperature distribution along the transmission line is known in analytic form. Exact and approximate solutions are listed in Table 2 for various typical temperature distributions. The approximations are especially useful where the transmission line loss is known in decibels and are quite accurate when the losses are low. Ususally only the first and second terms in the expansion are used. The third term can be examined as an indication of the error involved in the approximate calculation.

These solutions are satisfactory for most reference termination calibrations as long as the transmission line loss is low. A digital computer program has been developed for use where $T_L(x)$ is given in arbitrary form and the additional accuracy is required. This program is available for general use and can be specified by the JPL identification number TLH 5475000.

b. 90-Gc hot termination and computer program calibration. Best accuracy is usually obtained in calibrating a low-noise receiving system by using a cooled termination. However, in the event that larger changes in calibration temperatures are needed, which is usually the case for present day receivers operating at millimeter waves, e.g., 90-Gc, hot terminations are useful. Fig. 13 shows a diagram of the hot termination used with the 90-Gc millimeter radiometer (described in Sect. A1 of this report). Because the loss of the 90-Gc waveguide is high (0.422 db/in.), it was decided to calibrate the equivalent noise temperature from the actual temperature distribution along the waveguide using the digital computer program. Fig. 14 shows a graph of the measured transmission line temperature $T_L'(x)$ for a particular value of the termination temperature T. The temperature distribution $T_L(x)$ for an arbitrary value of T is proportional to $T_L(x)$ except that it must approach a straight line as T approaches T_0, the temperature of the transmission line at

Table 2. Tabulation of the theoretical equivalent noise temperatures of a termination attenuated by transmission lines with various combinations of thermal temperature distributions

Transmission line conditions	Unknown	Exact	Approximation (Valid if transmission line loss is less than approximately 0.5 db)
Transmission line with loss L and constant temperature distribution T_0	T'	$T + (T_0 - T)\left(1 - \dfrac{1}{L}\right)$	$T + (T_0 - T)\mathcal{L} - (T_0 - T)\mathcal{L}^2 + \cdots$
	$(T' - T_0)$	$(T - T_0)\left(\dfrac{1}{L}\right)$	$(T - T_0)(1 - \mathcal{L} + \mathcal{L}^2 + \cdots)$
Transmission line with loss L and linear temperature distribution from T to T_0	T'	$T + (T_0 - T)\left(1 - \dfrac{1 - \frac{1}{L}}{\mathcal{L}}\right)$	$T + \dfrac{(T_0 - T)}{2}\mathcal{L} - \dfrac{(T_0 - T)}{6}\mathcal{L}^2 + \cdots$
	$(T' - T_0)$	$(T - T_0)\left(\dfrac{1 - \frac{1}{L}}{\mathcal{L}}\right)$	$(T - T_0)\left(1 - \dfrac{1}{2}\mathcal{L} + \dfrac{1}{6}\mathcal{L}^2 + \cdots\right)$
Transmission line with loss L_1 and linear temperature distribution from T to T_0 for Sec. 1 and loss L_2 and constant temperature distribution T_0 for Sec. 2	T'	$T + (T_0 - T)\left[1 - \dfrac{\left(1 - \frac{1}{L_1}\right)\left(\frac{1}{L_2}\right)}{\mathcal{L}}\right]$	$T + (T_0 - T)\left(\dfrac{1}{2}\mathcal{L}_1 + \mathcal{L}_2\right)$ $- (T_0 - T)\left(\dfrac{1}{6}\mathcal{L}_1^2 + \dfrac{1}{2}\mathcal{L}_2^2 + \dfrac{1}{2}\mathcal{L}_1\mathcal{L}_2\right) + \cdots$
	$(T' - T_0)$	$(T - T_0)\left[\dfrac{\left(1 - \frac{1}{L_1}\right)\left(\frac{1}{L_2}\right)}{\mathcal{L}_1}\right]$	$(T - T_0)\left[1 - \left(\dfrac{1}{2}\mathcal{L}_1 + \mathcal{L}_2\right) + \left(\dfrac{1}{6}\mathcal{L}_1^2 + \dfrac{1}{2}\mathcal{L}_2^2 + \dfrac{1}{2}\mathcal{L}_1\mathcal{L}_2\right) + \cdots\right]$

$$T' = \frac{T + \dfrac{\mathcal{L}}{l}\displaystyle\int_0^l \exp\left(\frac{\mathcal{L}x}{l}\right) T_L(x)\,dx}{L}$$

for a transmission line with loss L independent of temperature, length l, and temperature distribution $T_L(x)$ where

$$\mathcal{L} = \frac{L(\text{db})}{10\log_{10} e} \simeq 0.2303L\,(\text{db})$$

T, T' = input and output equivalent noise temperatures
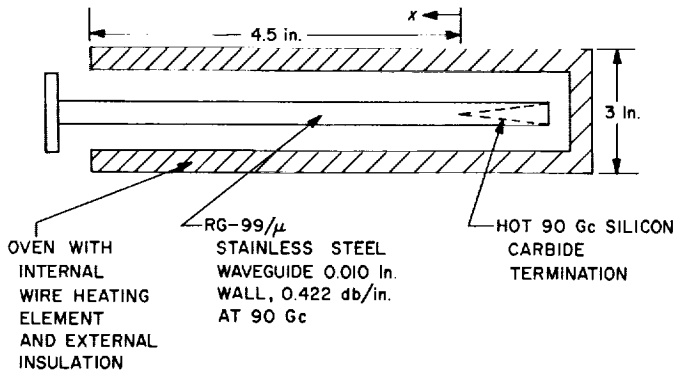
Fig. 13. 90-Gc hot termination in oven

DISTANCE FROM TERMINATION, x, in.
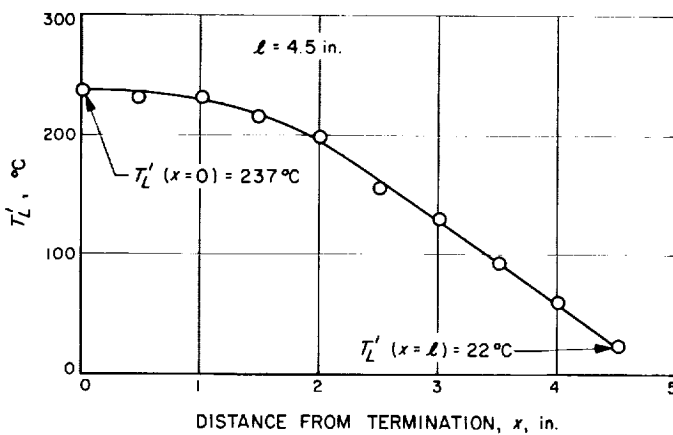
Fig. 14. Measured temperature distribution for 90-Gc radiometer hot termination transmission line

the unheated end. The expressions used to approximate these conditions are:

$$T_L(x) = T'_L(x)\left(\frac{l-x}{l}\frac{T}{T'_L(0)} + \frac{x}{l}\right)\left(\frac{T-T_0}{T'_L(0)-T_0}\right)$$
$$+ \frac{T-T'_L(0)}{T_0-T'_L(0)}\left[T-(T-T_0)\frac{x}{l}\right],$$

(2)

when

$$T \leqslant T'_L(0).$$

$$T_L(x) = T'_L(0)\left(\frac{l-x}{l}\frac{T}{T'_L(0)} + \frac{x}{l}\right),$$

when

$$T > T'_L(0).$$

If the temperature distribution were measured for each value of T, these expressions would not be needed. The programming is done so as to print out and graph T'

versus T for a given value of T_0 (294°K in this example) and $(T'-T_0)$ versus $(T-T_0)$ as shown in Table 3 and Figs. 15 and 16.

c. Calibration measurement error. The most critical measurement in the calibration of the equivalent noise temperature of a reference termination is the insertion loss of the transmission line. As an example, if the temperature distribution is constant (T_0) along a transmission line with loss L, the equivalent noise temperature is (Table 2)

$$T' \simeq T + 0.2303(T_0 - T)\,L\,(\text{db}) \tag{3}$$

This is accurate to better than 1% if the insertion loss $L(\text{db})$ is less than 0.1 db. The error in T' due to insertion loss measurement errors is (differentiating Eq. 3)

$$\Delta T' \simeq 0.2303\,(T_0 - T)\,\Delta L\,(\text{db}) \tag{4}$$

Table 3. 90-Gc hot termination computer program calibration printout

T_H	T'_H	$(T_H - T_0)$	$(T'_H - T_0)$
— 0.	31.158258	— 294.000000	— 262.841740
20.000000	49.384274	— 274.000000	— 244.615725
40.000000	67.559831	— 254.000000	— 226.440168
60.000000	85.684935	— 234.000000	— 208.315065
80.000000	103.759582	— 214.000000	— 190.240417
100.000000	121.783770	— 194.000000	— 172.216230
120.000000	139.757504	— 174.000000	— 154.242496
140.000000	157.680779	— 154.000000	— 136.319221
160.000000	175.553600	— 134.000000	— 118.446400
180.000000	193.375965	— 114.000000	— 100.624035
200.000000	211.147875	— 94.000000	— 82.852125
220.000000	228.869324	— 74.000000	— 65.130676
240.000000	246.540323	— 54.000000	— 47.459677
260.000000	264.160858	— 34.000000	— 29.839142
280.000000	281.730946	— 14.000000	— 12.269054
300.000000	299.250568	6.000000	5.250568
320.000000	316.719742	26.000000	22.719742
340.000000	334.138454	46.000000	40.138454
360.000000	351.506714	66.000000	57.506714
380.000000	368.824512	86.000000	74.824512
400.000000	386.091858	106.000000	92.091858
420.000000	403.308746	126.000000	109.308746
440.000000	420.475182	146.000000	126.475182
460.000000	437.591152	166.000000	143.591152
480.000000	454.656681	186.000000	160.656681
500.000000	471.671738	206.000000	177.671738
520.000000	488.074368	226.000000	194.074368
540.000000	504.005844	246.000000	210.005844
560.000000	519.937325	266.000000	225.937325
580.000000	535.868805	286.000000	241.868805

T_H is load input noise temperature,
T'_H is load output noise temperature, and
T_0 is temperature of termination for unheated end.
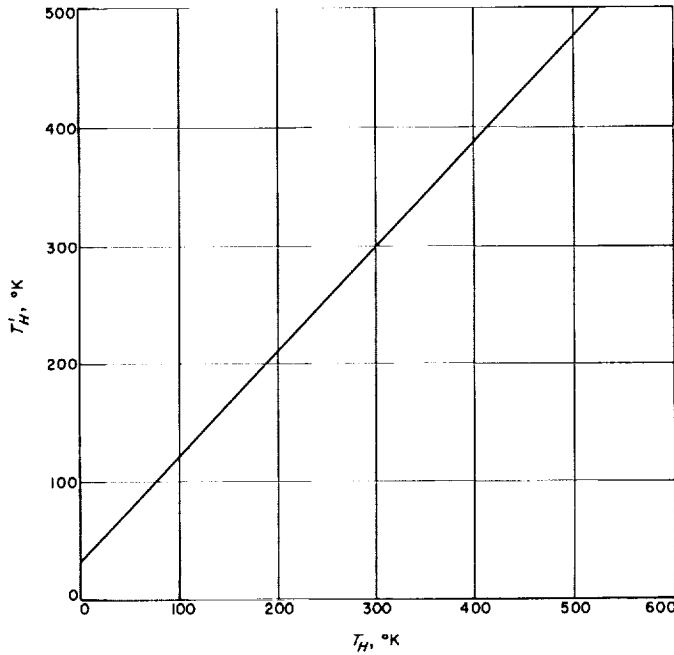
Fig. 15. 90-Gc hot load output noise temperature, T'_H, versus load input temperature, T_H; calibration from computer program of Table 3
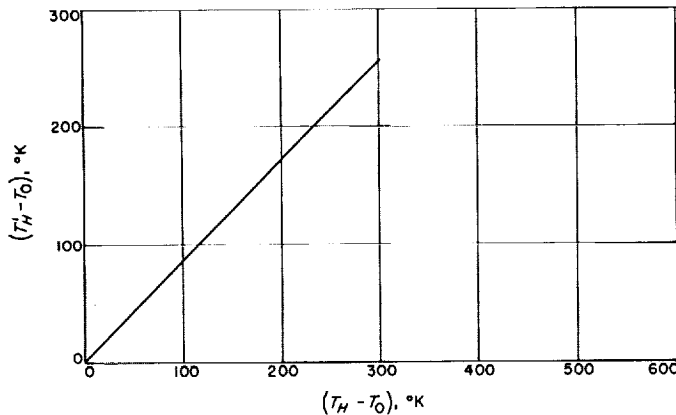
Fig. 16. 90-Gc hot load $(T' - T_0)$; calibration graph from computer program of Table 3

Equation (4) is plotted in Fig. 17 for a liquid helium and a liquid nitrogen cooled termination. To determine T' to an accuracy of 0.1°K for a liquid nitrogen cooled termination requires better than 0.002 db accuracy in the insertion loss measurements.

d. Calibration of equivalent noise temperature versus frequency. It is usually not fruitful to calibrate a reference termination at more than one frequency due to the narrow frequency range attainable in the VSWR matching.
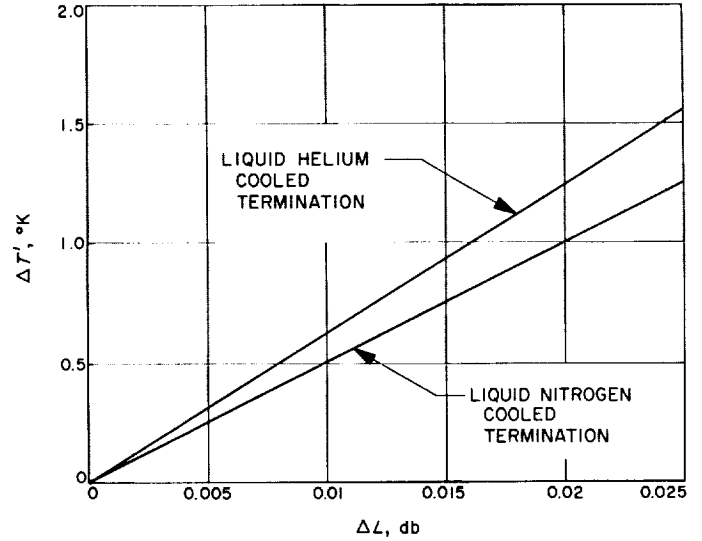
Fig. 17. Calibration error in a reference termination due to insertion loss measurement error (Eq. 4)

However, it is sometimes necessary to match a reference termination over a broad frequency range. After a termination is calibrated at a particular frequency, it is sometimes not worthwhile to remeasure the insertion loss at another frequency and recalibrate. Therefore, in these cases it is desirable to have an equivalent noise temperature calibration useful over a wide range of frequencies. Assuming small transmission line losses, loss L_1 in the first section of the transmission line with a linear temperature distribution from T to T_0, and loss L_2 in the second section at a constant temperature T_0, we have (Table 2):

$$T' \simeq T + 0.2303\,(T_0 - T)\left[\frac{L_1\,(\mathrm{db})}{2} + L_2\,(\mathrm{db})\right].$$
(5)

For a transmission line of length l with an over-all insertion loss $L_0(\mathrm{db})$ at frequency f_0, the insertion loss $L(\mathrm{db})$ at any other frequency f, neglecting dielectric loss is approximately (Ref. 5)

$$L\,(\mathrm{db}) \simeq L_0\,(\mathrm{db})\left(\frac{f}{f_0}\right)^{\frac{1}{2}}$$
(6)

Assuming a uniform transmission line,

$$L_1\,(\mathrm{db}) \simeq L\,(\mathrm{db})\,l_1/l = L_0\,(\mathrm{db})\,(f/f_0)^{\frac{1}{2}}\,l_1/l$$
(7)

$$L_2\,(\mathrm{db}) \simeq L\,(\mathrm{db})\,l_2/l = L_0\,(\mathrm{db})\,(f/f_0)^{\frac{1}{2}}\,l_2/l$$

Substituting Eqs. (7) into (5),

$$T' \simeq T + \frac{0.2303\,(T_0 - T)\,L_0\,(\text{db})}{l}\left(\frac{l_1}{2} + l_2\right)(f/f_0)^{\frac{1}{2}}$$

(8)

Equation (8) is the solution for the transmission line input equivalent noise temperature in terms of measurable quantities and the operating frequency.

A commercial (Model SP 9119, Maury Microwave Corporation) stainless steel coaxial transmission line and termination suitable for operation at cryogenic temperatures has been installed and operated in a polystyrene container at liquid nitrogen temperatures (Figs. 18 and 19). The 1 liter container has about a 3-hr operating life between refills. Although previous terminations of this type have been useful only over a very limited frequency range ($\simeq 1\%$), this termination will operate over a 1 Kmc to a 4 Kmc range with a VSWR less than 1.06 at liquid nitrogen temperatures.

For the transmission line assembly Serial 018, $l = 5$ in., $l_1 = 1.70$ in., $l_2 = 2.70$ in., $L_0 = 0.093$ db, $f_0 = 2295$ Mc, $T = 77.4°K$, and $T_0 = 294°K$. Substituting in Eq. (8)

$$T' - T + 0.0689\,f^{\frac{1}{2}}, (f \text{ in Mc})$$

(9)

Eq. (9), the equivalent noise temperature calibration for the termination is shown plotted in Fig. 20. At 2295 Mc, a 10% error in the evaluation of T' only results in 0.3°K
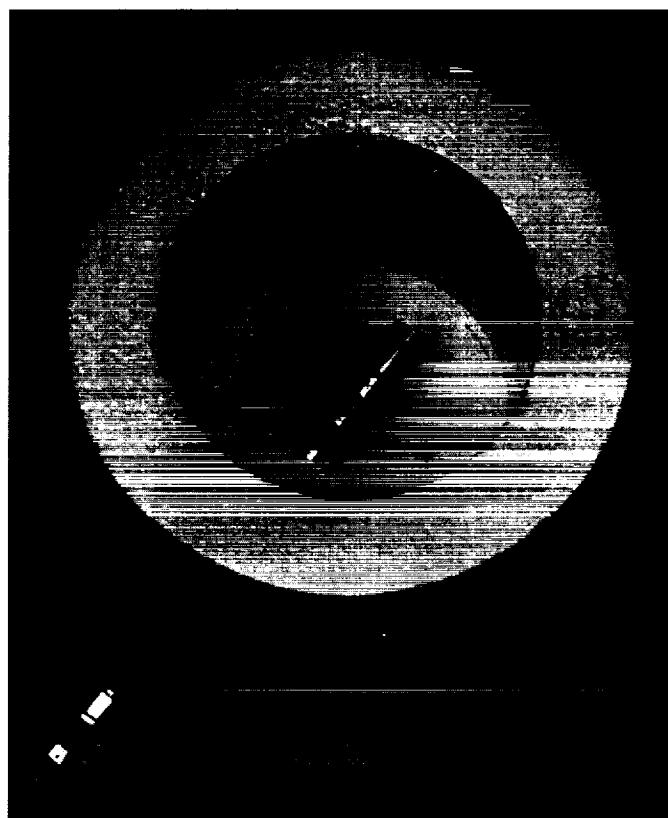
Fig. 19. Photo of liquid nitrogen cooled termination and container with N-connector output

Fig. 20. Calibration curve of equivalent input noise temperature T' versus frequency for transmission line assembly Ser. 018

error in T'. This again demonstrates the importance of keeping the transmission line losses low. The transmission line assembly was calibrated independently at 960 Mc with agreement in T' better than 0.1°K. The liquid nitrogen

Fig. 18. Diagrammatic view of liquid nitrogen cooled termination and container with N-connector output

level versus time in the polystyrene container with the coaxial transmission line and termination installed is shown in Fig. 21. The operating life of this combination is about 3 hr. The VSWR is not affected by the liquid level.

TIME AFTER FILL, hr
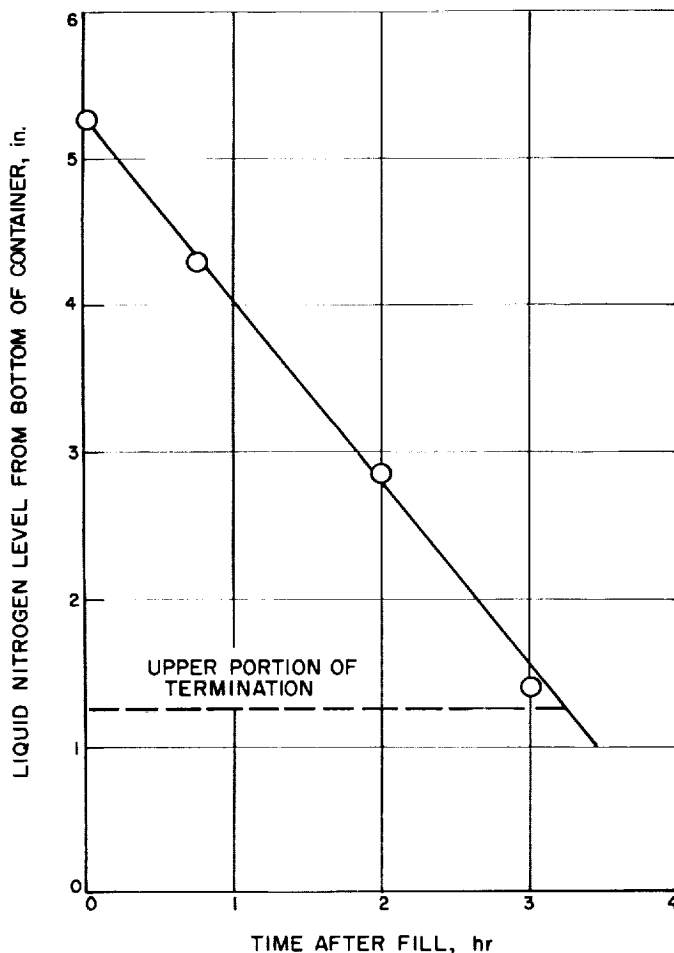
Fig. 21. Liquid nitrogen consumption for coaxial termination installed in polystyrene container

B. Optical Communications Components

J. Siddoway, H. Erpenbach, and C. Finnie

1. Lasers, *J. Siddoway*

a. Summary. Multiple-pass gain measurements at 1.15 and 0.6328 μ have been made with the traveling wave laser amplifier. Gain at the 1.15 μ wavelength is 7 db for

56 passes, but the amplifier is still lossy at 0.6328 μ. The latter spectral line is not suited for the present configuration with large gas volume.

b. Recent work. The optical system and gain measurement techniques for the traveling wave laser (TWL) amplifier were discussed in detail in Ref. 6. To review briefly, the TWL amplifier has a white absorption cell optical system which is the standard optical technique for obtaining long light paths in a gas. In this case, the gaseous medium has a negative absorption coefficient, and the gain of the system is determined by measuring the input power from the laser oscillator and the output power for a given number of traversals through the system. Gain measurements reported previously were for the 3.39 μ wavelength in He-Ne (40-db gain for 12 traversals). Gain curves which have since been obtained for the 1.15 μ wavelength are shown in Fig. 22. Curve A is the gain versus the number of traversals for an input power of 24 μw. Evidence of saturation effects is noticeable beyond 20 traversals. Curve B is for an input power of 72 μw with an abrupt change in the gain characteristics now at 12 traversals where saturation begins. The TWL amplifier is lossy at the 0.6328 μ wavelength. The measured loss per pass in the optics is $\simeq 1.5\%$, only partially compensated by the single-pass gain in the gas of less than 1%; for example, a typical power measurement is -3 db for 40 traversals.
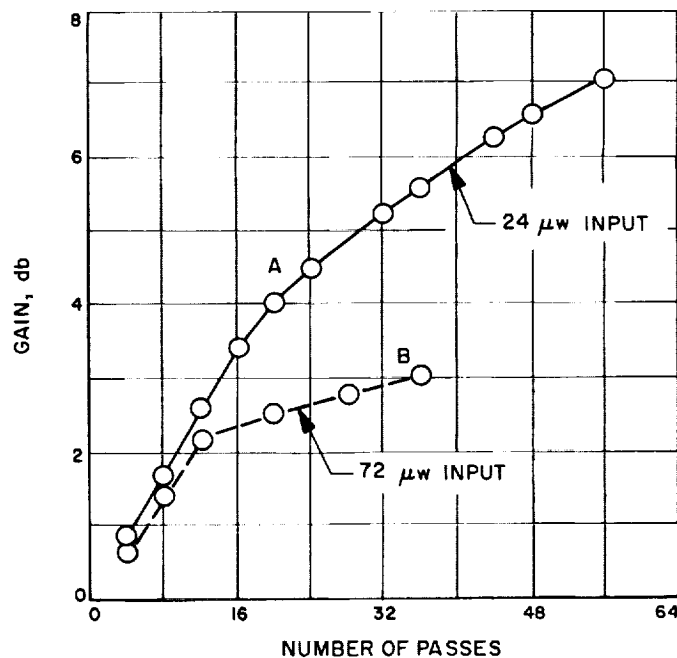
NUMBER OF PASSES

Fig. 22. Gain versus number of passes for 1.15 μ line in He-Ne

In analyzing the present performance of the TWL amplifier, one must consider the gain characteristics of the various wavelengths in this system and compare them to gain measurements of conventional laser tubes. In He-Ne laser systems, wall collisions are necessary in the pumping cycle of the atom and the gain/unit length of the gas depends upon the radius of the tube, roughly as $1/r$. Therefore, the advantages of multiple traversals through the gas may be overcome by the decrease in gain of the gas due to the larger volume, particularly for very high gain and low gain wavelengths. For example, the measured gain of the TWL amplifier at 3.39 μ is 40 db. In private communications, other researchers have reported single-pass gains of 50 db/m in small-bore (1.5-mm) tubes. Hence, for very high gain wavelengths a linear single-pass amplifier would be simpler and more efficient. For very low gain wavelengths, such as the 0.6328 μ line, where one might expect a net gain of 1 or 2%/pass, the TWL is limited to 2- or 3-db gain by the number of passes that can reasonably be obtained with the optics. The TWL amplifier is best suited for wavelengths such as the 1.15 μ spectral line. Single-pass gain of the gas is $\simeq$ 12%/m in conventional laser tubes with diameters comparable to the narrow dimension of the TWL amplifier. Presently, the net gain of the TWL amplifier is $\simeq$ 7%/pass; the amplifier is being modified with longitudinal spacers within the tube to reduce the volume and hence increase the gain and saturation level. Another possibility for a TWL amplifier configuration is an array of short small-bore tubes with the white cell optical system. This provides a fairly compact design and allows some spatial discrimination against noise (spontaneous emission) not possible with linear amplifiers.

2. Superconducting Bolometers,

H. Erpenbach and C. Finnie

a. Summary. Investigations of a superconducting bolometer have continued, using a signal from an electric heater for test purposes. In a normal application, the heat would come from absorbed RF, millimeter wave, infrared, or optical energy that the bolometer is intended to detect. Previous work was reported in the five preceding issues of this summary. The minimum detectable power quoted in Ref. 7 was in error and should read 1.5 $\times$ 10^{-7} w, not 1.4 $\times$ 10^{-13} w.

Recently, a thin film of tantalum was deposited on the end of a 0.25-in. D, 2-in. long sapphire rod. A nichrome heater ring was deposited 0.5 in. from one end around the rod circumference for raising the bolometer temperature from 4.2°K to 4.39°K, which is in the intermediate transition region when the bolometer operates. A small high-vacuum cryostat was fabricated for testing bolometers.

The new bolometer element has a folded path which was made by photochemical techniques. The minimum detectable signal is 6.9 $\times$ 10^{-8} w. Voltage-current characteristics are discussed in this summary; the geometry of heat flow has not yet been optimized for sensitivity.

b. Recent work. Mechanical masking techniques have limited all of the tantalum thin-film bolometer elements used in recent experimental work to approximately the same linear dimensions; i.e., 1000 A thick, 1 mil wide, and 125 mils long. Films of these dimensions have an electric resistance of approximately 15 Ω at room temperature and 1.4 Ω at 4.39°K, which is about the center of the transition region where the bolometer normally operates.

The low impedance of the bolometer element requires the use of a load matching transformer that could offer problems in future applications of the device. A bolometer with an impedance of at least 50 Ω in the center of the transition region would be desirable. We have recently fabricated and tested a new bolometer having 110 Ω at the operating temperature. Photochemical etching techniques were used to shape a multigrid pattern, a photograph of which is shown in Fig. 23. The tantalum thin film was deposited on the end of the cylindrical sapphire substrate without a mask. The film was coated with a metal etch resist and spun in a centrifuge radially 1 in. off center to 3000 rpm for 30 sec. Next, the etch resist was partially dried and spread out evenly to a thickness of about 3000 A, then prebaked at 120°C for 30 min. It was next mounted on an optical bench, and an image of the grid pattern was exposed onto the etch resist for about 5 sec. It was then placed in a developer which washed away the unexposed part of the grid, then baked another 30 min to drive out all of the solvents to insure complete polymerization of the etch resist. The final operation was to etch away the tantalum from which the etch resist had been removed, this by a 1- to 2-sec immersion into hydrofluoric acid. The physical characteristics of the completed element are

Area of grid	0.150 $\times$ 0.150 in.^{2}
Length of grid	6 in.
Width of tantalum	0.003 in.

Width of spacing	0.001 in.
Lines per inch	280
Normal resistance of element	1350 Ω at 25°C

Fig. 23. Folded path bolometer element

Resistance at center of transistor region	110 Ω at 4.39°K

Preliminary tests were performed on the bolometer. A noise equivalent signal was applied to the heater element, similar to testing described in Ref. 7. The noise equivalent signal was 9.6×10^{-8} w for a 150-cps noise bandwidth about a 1000-cps center frequency. The oscilloscope trace (Fig. 24) is the rectified noise and signal outputs of the bolometer amplifier. A filter time constant of 2 sec was used for this display. A transistorized, battery supplied, oscillator was used as the equivalent signal source.

Characteristics of the bolometer are plotted in Fig. 25. The long superconducting path of this bolometer (6 in. as compared to 3⁄16 in. of the previously reported design) has raised the resistance to 110 Ω at operating temperature. Approximately 10 Ω residual resistance exists in the superconducting state. The characteristics do not contain the negative resistative-thermal properties of the previous design. Rather, they are of the shape of the characteristic obtained previously when the temperature was raised by means of the dc heater.

The minimum detectable signal is only a factor of 1.6 lower than that of the previous low-impedance bolometer. The value of 1.4×10^{-13} w that appeared in Ref. 7 is incorrect. Changes need to be made in the geometry of the heat conduction path to achieve greater sensitivity.

BOLOMETER RECTIFIED NOISE

(1) ZERO BIAS RECEIVER NOISE, i.e., ELEMENT SUPERCONDUCTING
(2) BOLOMETER NOISE
(3) BOLOMETER NOISE PLUS 9.6×10^{-8} w, 1000 cps HEATER SIGNAL

Fig. 24. Bolometer rectified noise
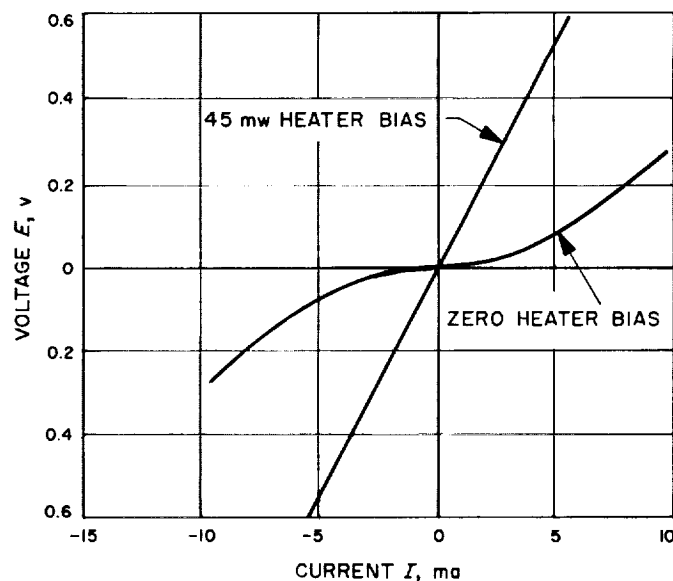
Fig. 25. Bolometer element voltage-current characteristics

C. Antennas for Space Communications

P. D. Potter and A. Ludwig

1. Antenna Feed Research; Spherical Wave Functions, *P. D. Potter*

Summary. The theory of spherical electromagnetic waves was reviewed in Ref. 8. In that report, it was shown that spherical wave theory could be utilized to predict ultimate limitations on paraboloidal antenna performance, imposed fundamentally by the laws of electromagnetic propagation (Maxwell's equations).

In this summary, the preliminary results of this study are presented. Curves of maximum possible aperture efficiency, lowest zenith noise temperature and optimized antenna figure of merit are presented as a function of the paraboloid diameter in wavelengths. One of the objectives, stated in Ref. 8 for the spherical wave study is to provide an analytical synthesis technique for non-optical Cassegrain subreflectors with increased performance. This will be the subject of a later reporting.

b. Radiation pattern formulation. It was shown in Ref. 8, that an arbitrary desired feed system radiation pattern can be expanded in a series of *TE* and *TM* waves. This series is, in general, infinite and must be truncated because of the relationship between wave order and antenna physical size. With the truncated series, however, it is possible to obtain by use of both *TE* and *TM* waves (or modes), a feed radiation pattern which produces a completely symmetrical paraboloid aperture distribution with zero aperture cross polarization and negligible phase error. In the far field, such a feed system has a radiation pattern $E(R, \psi, \xi)$ given by:

$$E(R, \psi, \xi) = \frac{Z_0 e^{j(\omega t - kR)}}{kR} (\sin \xi \, \mathbf{a}_\psi + \cos \xi \, \mathbf{a}_\xi)$$

$$\times \sum_{n=1}^{n=N} - (j)^n a_{TEn} \left[\frac{P_n^1(\psi)}{\sin \psi} + \frac{dP_n^1(\psi)}{d\psi} \right] \quad \text{(1a)}$$

where

$$a_{TEn} = \frac{-(-j)^n (2n+1)}{2Z_0 n^2 (n+1)^2} \int_0^\pi F(\psi) \left[\frac{P_n^1(\psi)}{\sin \psi} \right.$$

$$\left. + \frac{dP_n^1(\psi)}{d\psi} \right] \sin \psi \, d\psi \quad \text{(1b)*}$$

---

*This equation is similar to Eq. (16), p. 153, of SPS 37-24, Vol. IV, which had a typographical error in omitting the minus sign in front of the *j* within the parentheses.

where

R, ψ, ξ = the standard polar coordinate system coordinates

$\mathbf{a}_R, \mathbf{a}_\psi, \mathbf{a}_\xi$ = the associated unit vectors

k = free space propagation constant

ω = angular frequency

$Z_0 = 120\,\pi$ ohms

$F(\psi)$ = the desired "ideal," feed system polar pattern intensity

P_n^1 = associated Legendre polynomial of the first kind

If the field radiated by a feed system defined in Eqs. (1a) and (1b) is examined in the Fresnel region, a somewhat more complicated expression must be used. It is given by:

$$E(R, \psi, \xi) = Z_0 \sum_{n=1}^{n=N} a_{TEn}$$

$$\times \left\{ \begin{array}{l} \left\{ jh_n(\rho) \dfrac{P_n^1(\psi)}{\sin \psi} - \dfrac{1}{\rho} \dfrac{d}{d\rho} [\rho h_n(\rho)] \dfrac{dP_n^1(\psi)}{d\psi} \right\} \sin \xi \, \mathbf{a}_\psi \\[2mm] + \left\{ jh_n(\rho) \dfrac{dP_n^1(\psi)}{d\psi} - \dfrac{1}{\rho} \dfrac{d}{d\rho} [\rho h_n(\rho)] \dfrac{P_n^1(\psi)}{\sin \psi} \right\} \cos \xi \, \mathbf{a}_\xi \end{array} \right\}$$

$$\text{(2)}$$

where $\rho = kR$, and $h_n(\rho)$ is the spherical Hankel function. A useful asymptotic expansion for the Hankel functions has been developed by Stratton (Ref. 9).

$$h_n(\rho) = \frac{j^{(n+1)}}{\rho} e^{-j\rho} \left[P_{n+\frac{1}{2}}(\rho) - jQ_{n+\frac{1}{2}}(\rho) \right] \quad \text{(3)}$$

where

$$P_{n+\frac{1}{2}}(\rho) = 1 - \frac{n(n^2-1)(n+2)}{2^2 \cdot 2! \rho^2}$$

$$+ \frac{n(n^2-1)(n^2-4)(n^2-9)(n+4)}{2^4 \cdot 4! \rho^4} - \dots$$

$$\text{(4a)}$$

$$Q_{n+\frac{1}{2}}(\rho) = \frac{n(n+1)}{2 \cdot 1! \rho}$$

$$- \frac{n(n^2-1)(n^2-4)(n+3)}{2^3 \cdot 3! \rho^3} + \dots \quad \text{(4b)}$$

It should be noted that the series in Eqs. (4a) and (4b) are truncated, the number of terms being dependent on the order n. Further, for $\rho \geq n^2$, these series converge rapidly and only the leading terms are significant. Also, it can be readily shown that

$$\frac{1}{\rho}\frac{d}{d\rho}\left[\rho h_n(\rho)\right] \approx - jh_n(\rho), \qquad \rho \gg 2n \qquad (5)$$

$$|h_n(\rho)| \approx \frac{1}{\rho}, \qquad \rho \gg 2n \qquad (6)$$

Combining Eqs. (2), (5) and (6), one has

$$\mathbf{E}(R, \psi, \xi) \approx \frac{Z_0 e^{j(\omega t - kR)}}{kR}(\sin \xi\, \mathbf{a}_\psi + \cos \xi\, \mathbf{a}_\xi)$$

$$\times \sum_{n=1}^{n=N} -(j)^n e^{j\delta_n}\mathbf{a}_{TEn}\left[\frac{P_n^1(\psi)}{\sin \psi} + \frac{dP_n^1(\psi)}{d\psi}\right] \qquad (7)$$

$$\text{for } kR \geq N^2$$

where the mode phase error δ_n is given by

$$\delta_n = -\tan^{-1}\frac{Q_{n+\frac{1}{2}}(kR)}{P_{n+\frac{1}{2}}(kR)} \approx -\tan^{-1}\frac{\dfrac{n^2}{2\rho}}{1 - \dfrac{n^4}{8\rho^2}}$$

$$\qquad (8)$$

$$\delta_n \approx -\tan^{-1}\frac{1}{2}\frac{n^2}{\rho}\left[1 + \frac{1}{8}\left(\frac{n^2}{\rho}\right)^2\right]$$

For $n^2 = kR$, $\delta_n < 29°$. Since this phase error is comparable to the usual "far field" condition of $\lambda/16$ for antenna apertures, it will be assumed that the maximum usable N is given by

$$N = (kf)^{\frac{1}{2}} = \left(\frac{2\pi f}{D}\right)^{\frac{1}{2}}\left(\frac{D}{\lambda}\right)^{\frac{1}{2}} \qquad (9)$$

where D is the paraboloid diameter, f the paraboloid focal length, and λ the operating wavelength. If the "ideal"
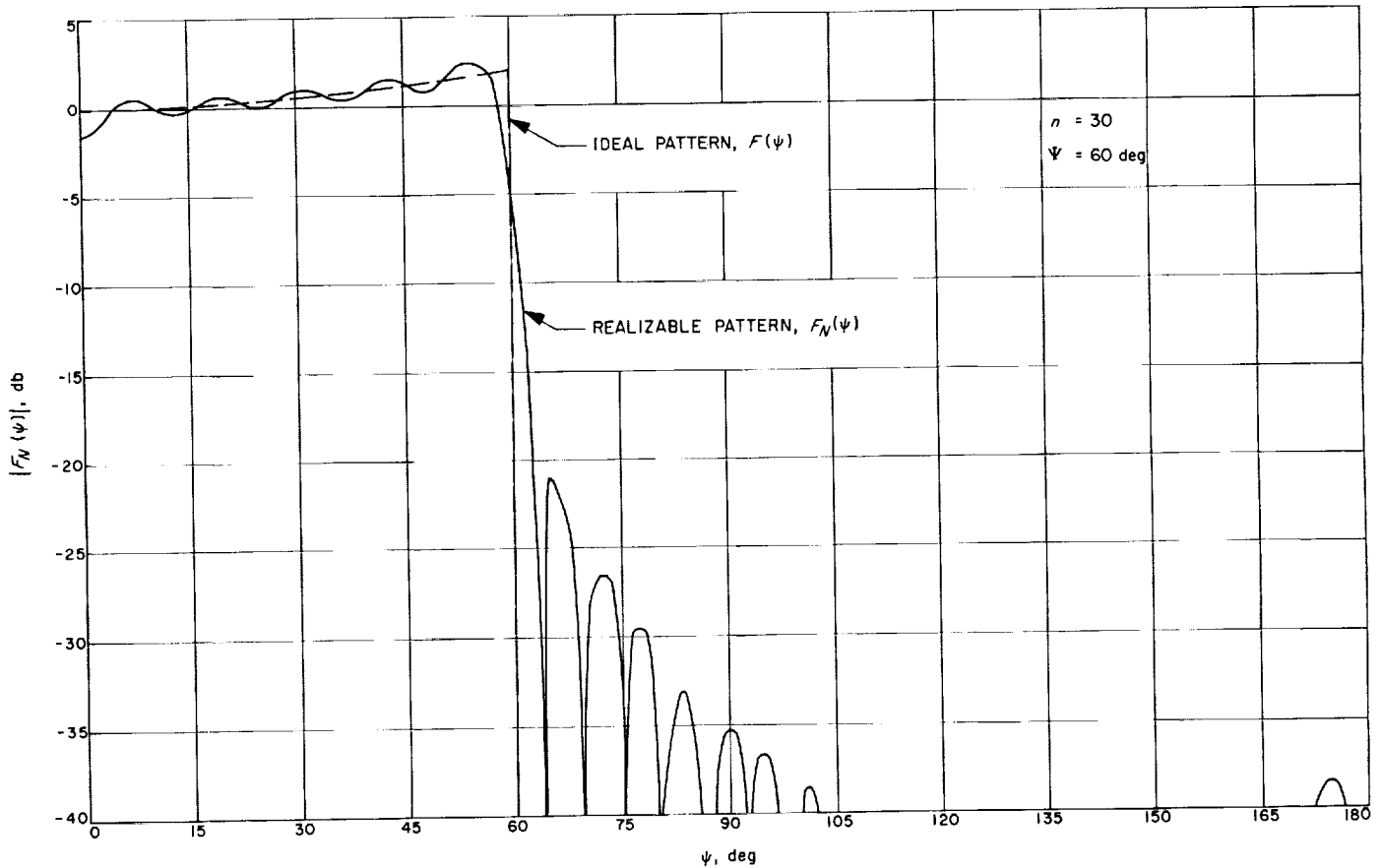
Fig. 26. Optimum feed pattern, $N = 30$

feed system polar radiation pattern, $F(\Psi)$ (see Eq. 1b) is picked to provide uniform paraboloid aperture illumination with no spillover, Eqs. (1a), (1b) and (9) may be used to calculate the "best" realizable feed system radiation pattern for a paraboloid of given wavelength diameter. Fig. 26 shows such a polar feed pattern for the case of $n = 30$. For an f/D ratio of 0.4235 (such as the Goldstone 85-ft antennas), this case requires, by Eq. (9), a minimum paraboloid diameter of 338 wavelengths. The almost uniform illumination and very low spillover is evident in Fig. 26.

c. Calculated performance limitations. Eqs. (1) have been numerically evaluated by machine computation up to $n = 30$, and used with Eq. (9) to develop optimum feed system radiation patterns as a function of paraboloid diameter in wavelengths. These patterns have, in turn, been numerically integrated by machine to calculate zenith noise temperature, T_s, and aperture efficiency, η, versus antenna diameter in wavelengths. These two performance parameters were calculated for each case as a function of the paraboloid subtended half-angle, Ψ. For each case, a value of Ψ was picked which maximized a figure-of-merit, FM, where

$$FM \overset{\Delta}{=} \frac{\eta T_0}{T_0 + T_s} \tag{10}$$

and

$$T_s \overset{\Delta}{=} (240°K) \times (\text{rear hemisphere spillover}) \tag{11}$$

Fig. 27 shows the optimum paraboloid subtended half-angle, Ψ_{opt}, as a function of D/λ, for various values of T_0, the sum of the atmospheric, sky background and the receiver system noise temperatures. The case of $T_0 = 20°K$
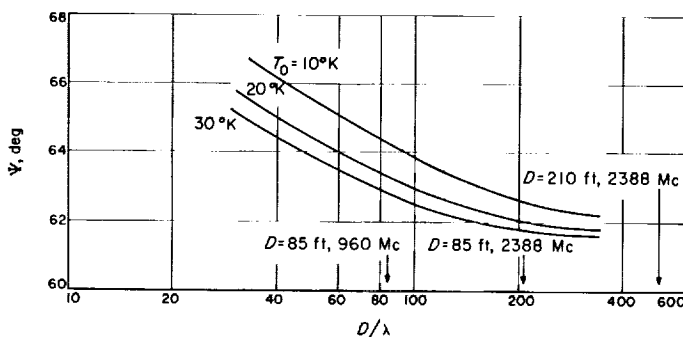
roughly corresponds to modern traveling wave maser receiving systems. Figs. 28, 29 and 30 show respectively the spillover noise contribution T_s; the aperture efficiency, η; and the relative figure of merit FM at Ψ_{opt} for various values of T_0. Fig. 30 predicts a difference of approximately 1.6 db in figure-of-merit between the best existing feed system (the Planetary Radar installation) and the performance this antenna could theoretically achieve.
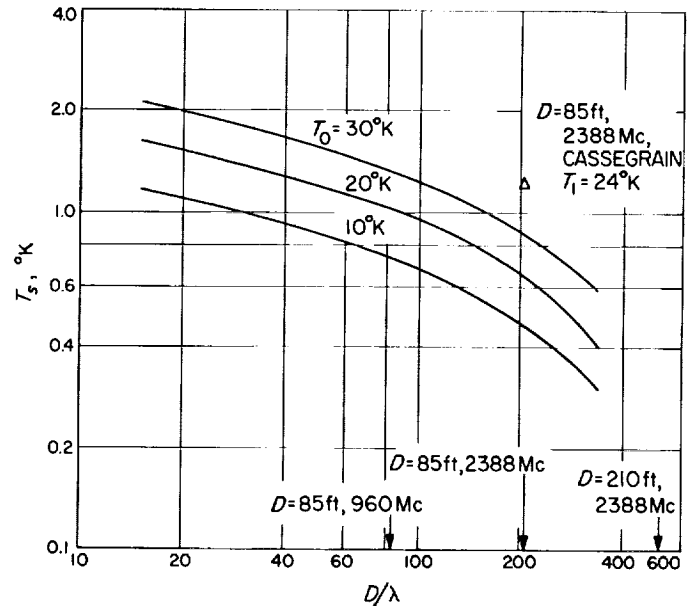
Fig. 28. Minimum feed system noise temperature contribution

Fig. 27. Optimum paraboloid subtended half-angle
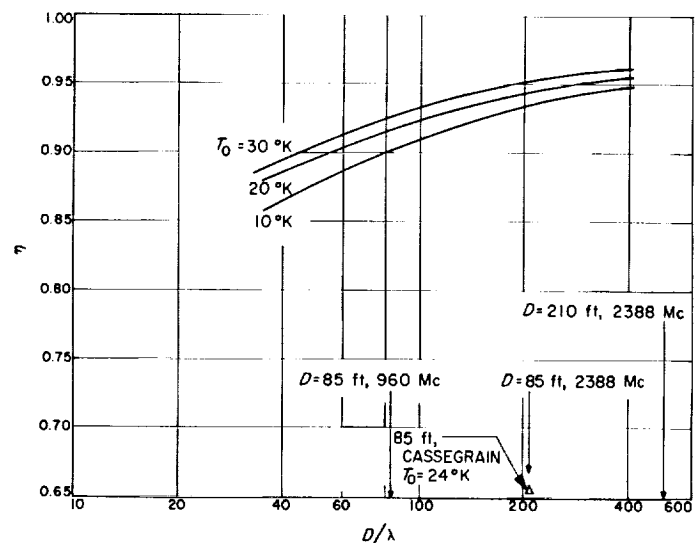
Fig. 29. Maximum feed system aperture efficiency
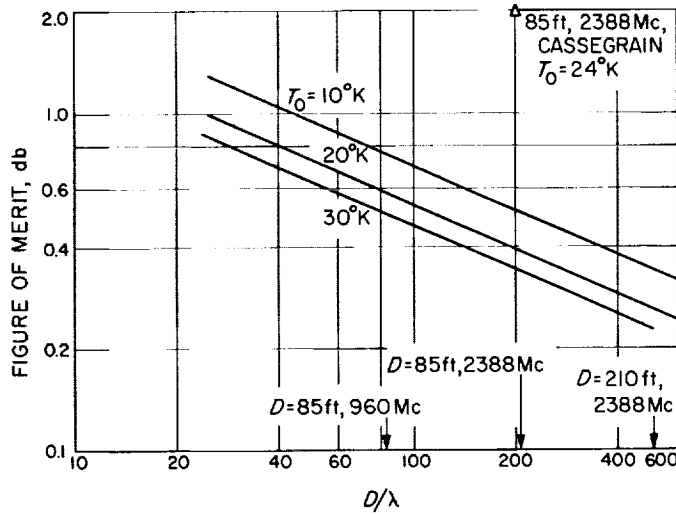
Fig. 30. Maximum feed system figure-of-merit

Although this study has predicted substantial (1 to 2 db) possible improvement in feed system design, practical, non-fundamental limitations such as quadripod and Cassegrain subreflector blockage have not yet been included in the analysis. It is intuitively expected that inclusion of these limitations will reduce the potential feed system improvement to 1 db or less.

2. Antenna Feed Efficiency, A. Ludwig

a. Summary. A computer program has been completed which evaluates the efficiency of a paraboloidal antenna as a function of its feed pattern characteristics. The program is intended as an engineering tool for use in feed system design and development. Losses due to blockage by the quadripod support[1], reflector surface roughness, and other causes associated with the main antenna structure are not considered. Total loss is factored into contributions due to spillover, cross-polarization, reflector illumination, and phase errors. Computed results are presented for a test case and experimental patterns of three feed systems.

b. Theory. The equations evaluated by the program are developed following the method used by Silver (Ref. 10, Chap. 12) generalized to include noncircularly symmetric feed patterns with non-uniform phase distributions. Using the surface-current distribution method (Ref. 10, pp. 144–149), the field intensity in the secondary pattern

[1] Quadripod blockage is included in another program which calculates efficiency for circularly-symmetric, uniform-phase feed patterns (Ref. 11).

on the axis a distance R_0 from the focus may be shown to be the integral over the paraboloidal surface,

$$\mathbf{E}\,(R_0, 0, 0) = -\frac{j\omega\mu}{\pi R_0}\, e^{-jkR_0} \int_0^{2\pi}\int_0^\Psi \Big\{\, \mathbf{n}\times\mathbf{H}_0$$

$$- \big[(\mathbf{n}\times\mathbf{H}_0)\cdot\mathbf{i}_z\big]\,\mathbf{i}_z \,\Big\}\; e^{-jk\rho\,\cos\psi}\,\rho^2\,\sin\frac{\psi}{2}\,d\psi\,d\xi \qquad (1)$$

where the coordinate system is shown in Fig. 31

$j = \sqrt{-1}$

ω = frequency, radians/sec

μ = magnetic permeability

$\mathbf{n}$ = unit vector normal to the reflector surface

$\mathbf{H}_0$ = magnetic field pattern of the feed, a complex vector function of ρ,ξ,ψ

$\mathbf{i}_z$ = unit vector in the z-direction (in general $\mathbf{i}$ will be used as a unit vector in the direction of its subscript).

Assuming that the reflector is in the far-field $(\rho \gg \lambda)$ of the feed radiation pattern,

$$\mathbf{H}_0\,(\rho, \xi, \psi) = \left(\frac{\epsilon}{\mu}\right)^{1/2}\big[\mathbf{i}_\rho\times\mathbf{E}_0\,(\rho, \xi, \psi)\big] \qquad (2)$$
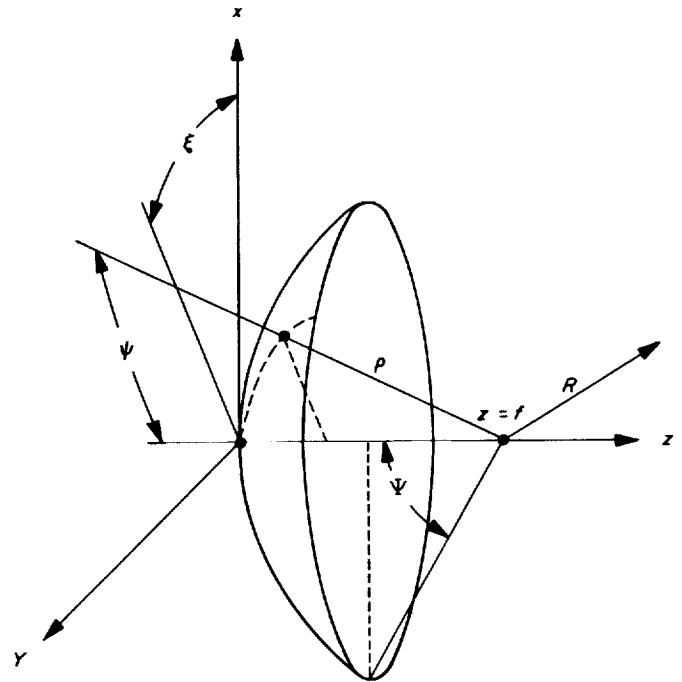
Fig. 31. Coordinate system

where $\mathbf{E}_0\ (\rho, \xi, \psi)$ is the electric field pattern of the feed. A completely general far-field expansion for any possible $\mathbf{E}_0$ in a source-free region is given by

$$\mathbf{E}_0\ (\rho, \xi, \psi) = \frac{e^{-jk\rho}}{\rho} \sum_{m=0}^{\infty} A_m\ (\psi) \sin\ (m\xi + \alpha_m)\, \mathbf{i}_\psi$$

$$+\ B_m\ (\psi)\ \cos\ (m\xi + \beta_m)\, \mathbf{i}_\xi \qquad (3)$$

where $A_m(\psi)$ and $B_m(\psi)$ are complex functions of ψ only, and α_m and β_m are constants. It might be pointed out that experimentally determining the coefficients $A_m(\psi)$ and $B_m(\psi)$ can be a difficult task: Rigorously, it is necessary to take a set of patterns of both E_ψ and E_ξ over a conical-rim path, $\psi = $ constant, and expand the resulting patterns in Fourier series. This must be done for as many values of ψ as necessary to obtain smooth functions. In practice, the possible values of m may be sharply restricted due to symmetry considerations, and it is only necessary to look for these coefficients. For example, feed horns with complete circular symmetry, excited with the dominant or any $m = 1$ cylindrical waveguide mode, will have radiation patterns containing only $m = 1$ terms[2]. Let

$$A_m\ (\psi) = |\,A_m\ (\psi)\,|\ e^{j\Phi_{Am}(\psi)} \qquad (4)$$

$$B_m\ (\psi) = |\,B_m\ (\psi)\,|\ e^{j\Phi_{Bm}(\psi)} \qquad (5)$$

If it is known that $A_m(\psi) = B_m(\psi) = 0$ for $m \neq 1$, then, by the usual definitions, $A_1(\psi)$ and $B_1(\psi)$ are respectively the E-Plane and H-Plane amplitude patterns and $\Phi_{A1}\ (\psi)$ and $\Phi_{B1}\ (\psi)$ are respectively the E-Plane and H-Plane phase patterns. For the paraboloidal reflector

$$\mathbf{n} = -\cos\left(\frac{\psi}{2}\right)\mathbf{i}_\rho + \sin\left(\frac{\psi}{2}\right)\mathbf{i}_\psi \qquad (6)$$

Using Eqs. (2) and (6) it is found that

$$\mathbf{n} \times \mathbf{H}_0 = \left(\frac{\epsilon}{\mu}\right)^{\!1/2} \left\{ [E_{0\psi} \cos \xi - E_{0\xi} \sin \xi]\ \cos\left(\frac{\psi}{2}\right)\mathbf{i}_x \right.$$

$$\left. +\ [E_{0\psi} \sin \xi + E_{0\xi} \cos \xi]\ \cos\left(\frac{\psi}{2}\right)\mathbf{i}_y + E_{0\psi} \sin\left(\frac{\psi}{2}\right)\mathbf{i}_z \right\}$$

$$(7)$$

[2]This is not true for rectangular feed horns, which in general will produce $m \neq 1$ terms.

Combining Eqs. (3) and (7) and using Eq. (1),

$$\mathbf{E}\ (R_0, 0, 0) = -\frac{j\omega\mu}{\pi R_0}\, e^{-jkR_0} \left(\frac{\epsilon}{\mu}\right)^{\!1/2} \sum_{m=0}^{\infty} \mathbf{I}_m \qquad (8)$$

where

$$\mathbf{I}_m = \int_0^{2\pi}\!\!\int_0^\Psi \left\{ \left[A_m\ (\psi) \sin\ (m\xi + \alpha_m) \cos \xi \right.\right.$$

$$-\ B_m\ (\psi) \cos\ (m\xi + \beta_m) \sin \xi \big]\ \mathbf{i}_x$$

$$+\ \big[A_m\ (\psi) \sin\ (m\xi + \alpha_m) \sin \xi$$

$$\left. +\ B_m\ (\psi) \cos\ (m\xi + \beta_m) \cos \xi \big]\ \mathbf{i}_y \right\}$$

$$\times \cos\left(\frac{\psi}{2}\right) e^{-jk\rho\ (1 + \cos \psi)}\, \rho \sin\left(\frac{\psi}{2}\right) d\psi\, d\xi \qquad (9)$$

Due to the orthogonality property of the sines and cosines, the integral over ξ is zero except for $m = 1$.

$$\mathbf{I}_m = 0 \qquad m \neq 1 \qquad (10)$$

$$\mathbf{I}_1 = \pi \int_0^\Psi \left\{\ [A_1\ (\psi) \sin \alpha_1 + B_1\ (\psi) \sin \beta_1]\ \mathbf{i}_x \right.$$

$$\left. +\ [A_1\ (\psi) \cos \alpha_1 + B_1\ (\psi) \cos \beta_1]\ \mathbf{i}_y \right\}$$

$$\times\ e^{-jk\rho\ (1 + \cos \psi)}\, \rho \sin\left(\frac{\psi}{2}\right) \cos\left(\frac{\psi}{2}\right) d\psi \qquad (11)$$

Thus components of the feed pattern, Eq. (3), with rotational variation of $m \neq 1$, fall in exactly the same category as cross-polarization; they contribute to total radiated energy and do not contribute to gain.

It is now assumed that there is a plane of symmetry $\xi = $ constant which divides the $m = 1$ contribution of $\mathbf{E}_0(\rho, \xi, \psi)$ into right- and left-hand mirror images. The coordinate system is now defined such that this becomes the $y - z$ plane ($\xi = 90°$). By symmetry $E_{0x} = 0$ in this plane, and it can be seen by Eq. (3) that $\alpha_1 = \beta_1 = 0$. Introducing the relationship for the paraboloid,

$$\rho = \frac{2f}{1 + \cos \psi} = f \sec^2\left(\frac{\psi}{2}\right) \qquad (12)$$

and using Eqs. (10) and (11) in Eq. (8),

$$E(R_0, 0, 0) = -i_y \frac{j\omega\mu f}{R_0} e^{-jk (R_0 + 2f)} \left(\frac{\epsilon}{\mu}\right)^{1/2}$$

$$\times \int_0^\Psi [A_1(\psi) + B_1(\psi)] \tan\left(\frac{\psi}{2}\right) d\psi \tag{13}$$

Gain G is defined as

$$G = \frac{P(0, 0)}{P_T/4\pi} \tag{14}$$

where

$$P(0, 0) = \text{power/unit solid angle on axis}$$

$$= \frac{1}{2} R_0^2 \left(\frac{\epsilon}{\mu}\right)^{1/2} |E(R_0, 0, 0)|^2 \tag{15}$$

and total radiated power

$$P_T = \frac{1}{2} \left(\frac{\epsilon}{\mu}\right)^{1/2} \int_0^{2\pi} \int_0^\pi |E_0(\rho, \xi, \psi)|^2 \rho^2 \sin\psi \, d\psi \, d\xi \tag{16}$$

using Eq. (3) in Eq. (16) and integrating over ξ,

$$P_T = \frac{\pi}{2} \left(\frac{\epsilon}{\mu}\right)^{1/2} \int_0^\pi \sum_{m=0}^\infty \left\{ |A_m(\psi)|^2 + |B_m(\psi)|^2 \right\} \sin\psi \, d\psi \tag{17}$$

using Eq. (13) in Eq. (15), and combining with Eq. (17), the gain, Eq. (14) becomes

$$G = 4\omega^2\epsilon\mu f^2 \frac{\left| \int_0^\Psi [A_1(\psi) + B_1(\psi)] \tan\left(\frac{\psi}{2}\right) d\psi \right|^2}{\int_0^\pi \left\{ \sum_{m=0}^\infty |A_m(\psi)|^2 + |B_m(\psi)|^2 \right\} \sin\psi \, d\psi} \tag{18}$$

Using the paraboloid relationship

$$f = \frac{D}{4} \cot\left(\frac{\Psi}{2}\right), \tag{19}$$

the relationship

$$\omega^2\epsilon\mu = \frac{\omega^2}{c^2} = \left(\frac{2\pi}{\lambda}\right)^2, \tag{20}$$

and the definition of antenna efficiency, (based on an ideal case of an aperture field of uniform amplitude, phase and polarization)

$$\eta = \frac{G}{G_0} = \frac{G}{(\pi D/\lambda)^2} \tag{21}$$

we obtain

$$\eta = \cot^2\left(\frac{\Psi}{2}\right) \frac{\left| \int_0^\Psi [A_1(\psi) + B_1(\psi)] \tan\left(\frac{\psi}{2}\right) d\psi \right|^2}{\int_0^\pi \left\{ \sum_{m=0}^\infty |A_m(\psi)|^2 + |B_m(\psi)|^2 \right\} \sin\psi \, d\psi} \tag{22}$$

For purposes of feed system design, it is convenient to factor Eq. (22) into several parts, thereby isolating causes of loss, and indicating which areas are most in need of improvement.

The denominator of Eq. (22), representing total radiated power, may be considered as the sum of three components; ϵ_s, representing energy spilling past the edge of the paraboloid,

$$\epsilon_s = \int_\Psi^\pi \left\{ \sum_{m=0}^\infty |A_m(\psi)|^2 + |B_m(\psi)|^2 \right\} \sin\psi \, d\psi \tag{23}$$

$\epsilon_{m \neq 1}$, energy striking the paraboloid but contained in $m \neq 1$ terms,

$$\epsilon_{m \neq 1} = \int_0^\Psi \left\{ \sum_{\substack{m=0 \\ m \neq 1}}^\infty |A_m(\psi)|^2 + |B_m(\psi)|^2 \right\} \sin\psi \, d\psi \tag{24}$$

and ϵ_1, energy striking the paraboloid contained in the $m = 1$ term,

$$\epsilon_1 = \int_0^\Psi \left\{ |A_1(\psi)|^2 + |B_1(\psi)|^2 \right\} \sin\psi \, d\psi \tag{25}$$

Eq. (25) may be written as

$$\epsilon_1 = \frac{1}{2} \int_0^\Psi \left[|A_1(\psi)| + |B_1(\psi)| \right]^2 \sin\psi \, d\psi$$

$$+ \frac{1}{2} \int_0^\Psi \left[|A_1(\psi)| - |B_1(\psi)| \right]^2 \sin\psi \, d\psi \tag{26}$$

Taking the $m = 1$ term only from Eq. (3), with $\alpha_1 = \beta_1 = 0$, and applying Eq. (7), it can be shown that the normal polarization component is:

$$(\mathbf{n} \times \mathbf{H}_0) \cdot \mathbf{i}_y = \left(\frac{\varepsilon}{\mu}\right)^{\frac{1}{2}} \cos\left(\frac{\psi}{2}\right) \left\{ \frac{1}{2}\left[A_1(\psi) + B_1(\psi) \right] \right.$$

$$\left. - \frac{1}{2}\left[A_1(\psi) - B_1(\psi) \right] \cos 2\xi \right\}, \quad (27)$$

and the cross-polarization component is:

$$(\mathbf{n} \times \mathbf{H}_0) \cdot \mathbf{i}_x = \left(\frac{\varepsilon}{\mu}\right)^{\frac{1}{2}} \cos\left(\frac{\psi}{2}\right) \left\{ \frac{1}{2}\left[A_1(\psi) \right. \right.$$

$$\left. \left. - B_1(\psi) \right] \sin 2\xi \right\}, \quad (28)$$

Using Eqs. (4) and (5),

$$|(\mathbf{n} \times \mathbf{H}_0) \cdot \mathbf{i}_x|^2 = \left(\frac{\varepsilon}{\mu}\right) \cos^2\left(\frac{\psi}{\mu}\right) \left\{ \frac{1}{2}\left[|A_1(\psi)| - |B_1(\psi)|\right]^2 \right.$$

$$\left. + 2|A_1(\psi)||B_1(\psi)| \sin^2\left[\frac{\Phi_{A1}(\psi) - \Phi_{B1}(\psi)}{2}\right] \right\} \sin 2\xi, \quad (29)$$

so cross-polarization from $m = 1$ terms arises due to: (a) difference in the E- and H-Plane amplitude patterns, and (b) difference in the E- and H-Plane phase patterns. Calling the first and second terms of Eq. (26) ε_{1n} and ε_{1x} respectively, we see that ε_{1x} represents energy in the cross-polarization component of the field due to cause (a) above. Taking total radiated power as constant, it follows that $\varepsilon_s + \varepsilon_{m \neq 1} + \varepsilon_{1n} + \varepsilon_{1x} = \varepsilon_T$, a constant, where ε_T is the denominator of Eq. (22). Spillover loss η_s is defined as

$$\eta_s \equiv \frac{\varepsilon_T - \varepsilon_s}{\varepsilon_T} . \quad (31)$$

In order to conveniently retain the property that η equal the product of all the losses defined separately, we take $m \neq 1$ loss as

$$\eta_{m \neq 1} \equiv \frac{[\varepsilon_T - \varepsilon_s] - \varepsilon_{m \neq 1}}{[\varepsilon_T - \varepsilon_s]} \cong \frac{\varepsilon_T - \varepsilon_{m \neq 1}}{\varepsilon_T} \quad (32)$$

and cross-polarization loss due to difference in E- and H-Plane amplitude patterns (cross-polarization loss due to difference in E- and H-Plane phase patterns is assigned to the category of phase loss) as:

$$\eta_{m \neq 1} \equiv \frac{[\varepsilon_T - \varepsilon_s - \varepsilon_{m \neq 1}] - \varepsilon_{1x}}{[\varepsilon_T - \varepsilon_s - \varepsilon_{m \neq 1}]} \cong \frac{\varepsilon_T - \varepsilon_{1x}}{\varepsilon_T}. \quad (33)$$

The difference between the actual definitions of η_x and $\eta_{m \neq 1}$ and the equations they approximate is less than 3% for $\varepsilon_s + \varepsilon_{m \neq 1} + \varepsilon_{1x} \leq 30\%$ of ε_T, a condition normally satisfied by a reasonable feed system. Antenna efficiency, Eq. (22), may now be written as

$$\eta = \eta_s \cdot \eta_{m \neq 1} \cdot \eta_x \cot^2 \frac{\Psi}{2}$$

$$\times \frac{\left|\int_0^\Psi [A_1(\psi) + B_1(\psi)] \tan\left(\frac{\psi}{2}\right) d\psi\right|^2}{\frac{1}{2}\int_0^\Psi [|A_1(\psi)| + |B_1(\psi)|]^2 \sin\psi \, d\psi}. \quad (34)$$

Phase loss is defined as the ratio of η with phase error present to η in the case $\Phi_{A1}(\psi) = \Phi_{B1}(\psi) = $ constant; the only term in Eq. (34) dependent on phase is the numerator, and we have

$$\eta_p = \frac{\left|\int_0^\Psi [A_1(\psi) + B_1(\psi)] \tan\left(\frac{\psi}{2}\right) d\psi\right|^2}{\left[\int_0^\Psi [|A_1(\psi)| + |B_1(\psi)|] \tan\left(\frac{\psi}{2}\right) d\psi\right]^2} \quad (35)$$

the remaining factor is defined as illumination loss,

$$\eta_I = \cot^2 \frac{\Psi}{2} \frac{\left[\int_0^\Psi [|A_1(\psi)| + |B_1(\psi)|] \tan\left(\frac{\psi}{2}\right) d\psi\right]^2}{\frac{1}{2}\int_0^\Psi [|A_1(\psi)| + |B_1(\psi)|]^2 \sin\psi \, d\psi} \quad (36)$$

η_I is a function of the average of the $m = 1$ E- and H-Plane amplitudes only, and represents loss due to deviations of the average from the ideal case,

$$\frac{1}{2}[|A_1(\psi)| + |B_1(\psi)|] = \sec^2\left(\frac{\psi}{2}\right), \quad (37)$$

which yields uniform amplitude in the aperture after optical reflection from the paraboloid. Overall efficiency is then

$$\eta = \eta_s \cdot \eta_{m \neq 1} \cdot \eta_x \cdot \eta_p \cdot \eta_I . \quad (38)$$

c. Computer program. Our current work on feed design deals exclusively with feedhorns of complete circular symmetry excited with $m = 1$ modes only. Therefore, it has been assumed that $\varepsilon_{m \neq 1} = 0$ (Eq. 24), and $\eta_{m \neq 1}$ is not computed by the program in its current form.

The program accepts E- and H-Plane magnitude (in decibels) and phase (in degrees) inputed in 2-deg increments of ψ, and linearly interpolates this data to 0.2-deg increments. Four integrals are evaluated using trapazoidal integration over the 0.2-deg increment, and manipulated

Fig. 32. E- and H-Plane amplitude and
phase patterns

to output η, η_s, η_x, η_p, and η_l (in percent) in 2 deg increments of ψ. Due to the fact that all quantities are a ratio of two similar integrals, interpolation and integration errors tend to cancel; the absolute error in the integration of Eq. (25) in a test case $A_1(\psi) = B_1(\psi) = \sec^2 \psi/2$ was 0.01% at $\psi = 26°$. Run time on the IBM 7094 is 15 sec/case.

d. Computed results. Two test cases (Figs. 32a, b) provide an interesting numerical example of the effect of unequal E- and H-Plane phase patterns (Eq. 29 and paragraph following it describe this effect). Both cases have ideal amplitude patterns, and linear phase errors in both planes. In Case 1, the edge of the beam ($\psi = 30$ deg) is 30 deg out of phase with the center in both planes. In Case 2, the edge of the beam is 15 deg and 45 deg out of phase with the center in the E- and H-Planes, respectively. At $\psi = 30$ deg, $\eta_p = 0.985$ (−0.13 db) for Case 1 and $\eta_p = 0.945$ (−0.48 db) for Case 2. So a *difference* in phase errors of 30 deg at the edge caused a considerably larger effect than 30 deg *average* phase error at the edge.

The three experimental cases taken were: (1) an ordinary dominant-mode conical horn; (2) a dual-mode horn (Ref. 12); and (3) a tri-mode horn using a TE_{12}^o mode generator (Ref. 13). Amplitude and phase patterns are shown in Figs. 33, 34, and 35. Overall efficiency versus ψ is shown in Fig. 36, and the breakdown into individual loss contributions, at peak efficiency in each case, is shown in Fig. 37. It is seen that phase and cross-polarization loss are virtually negligible, and that the two areas in need of improvement are spillover, and illumination loss. Spillover also contributes to antenna noise temperature, and for that reason becomes the primary target for attempted improvement.

e. Future work. Since Cassegrainian antenna systems are the intended application for these feeds, a realistic evaluation must consider the scattering pattern from the subreflector, rather than considering the feeds as focal point systems. To accomplish this, patterns will be input into a computer program designed by Dr. W. V. T. Rusch (Ref. 14), which calculates subreflector scattering, and efficiency will then be calculated for the scattered fields. This program has recently been modified to accept experimental input data, and is now ready for this application. Due to the finer detail in the scattered patterns, the
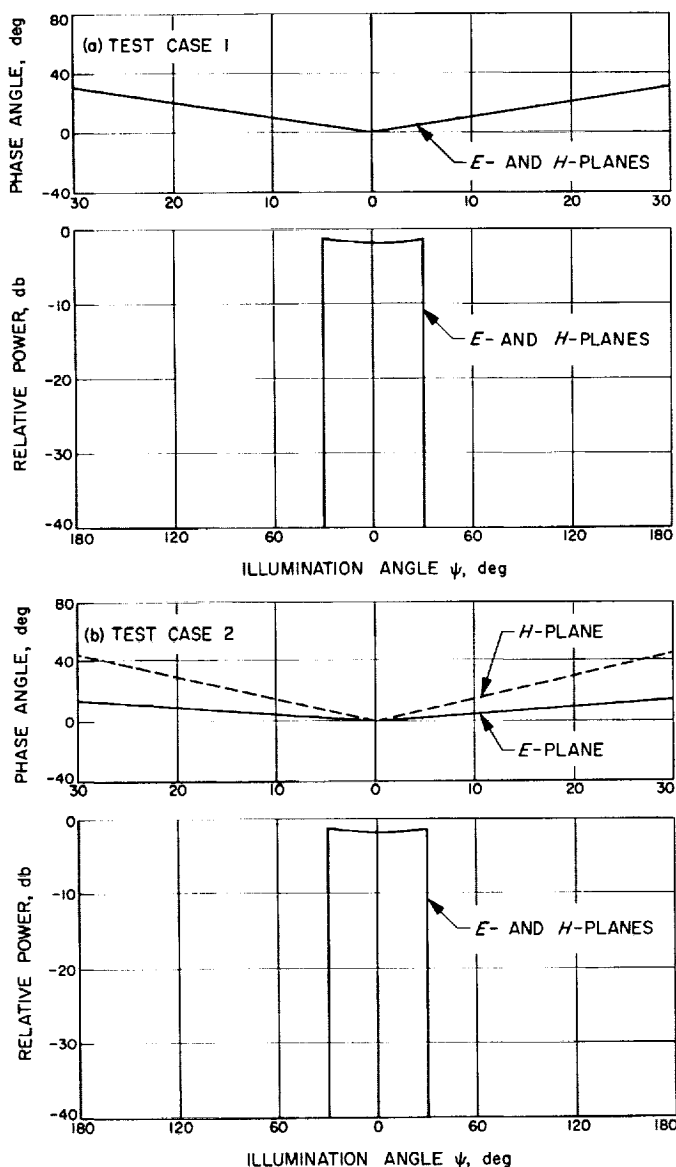
efficiency program will be modified to accept input data, and to output answers in 1-deg increments, rather than the present 2-deg increment.

This work is being correlated with the theoretical limitations calculated using spherical wave theory (Ref. 15, and preceding article) and the theoretical possibilities of cylindrical feedhorns (Ref. 16). This effort will hopefully provide an evaluation of the potential of feedhorns, compared to ultimate limitations on any type of feed of a given size, and an evaluation of how far we have come in realizing this potential.
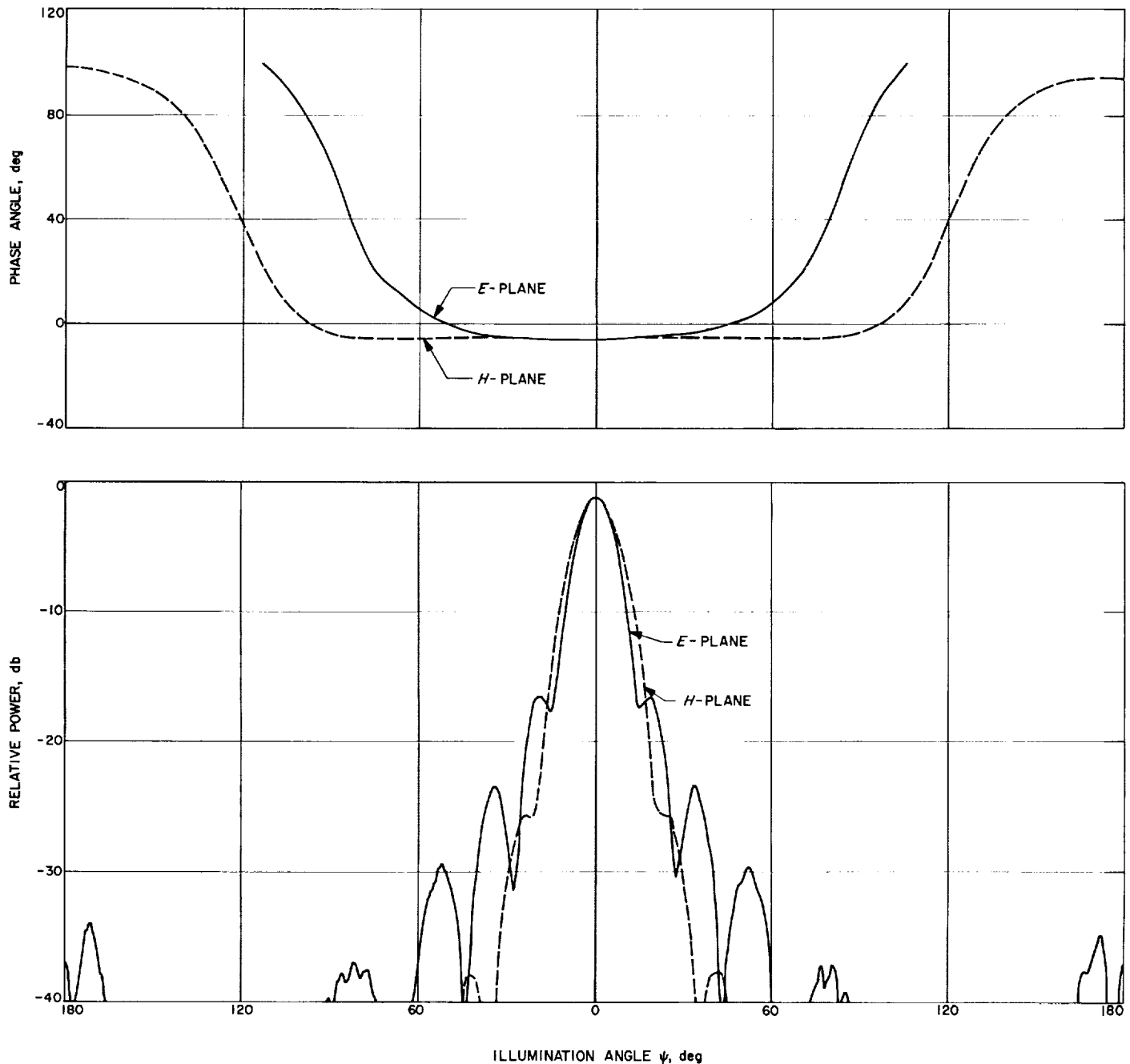
Fig. 33. Dominant-mode horn, E- and H-Plane amplitude and phase patterns
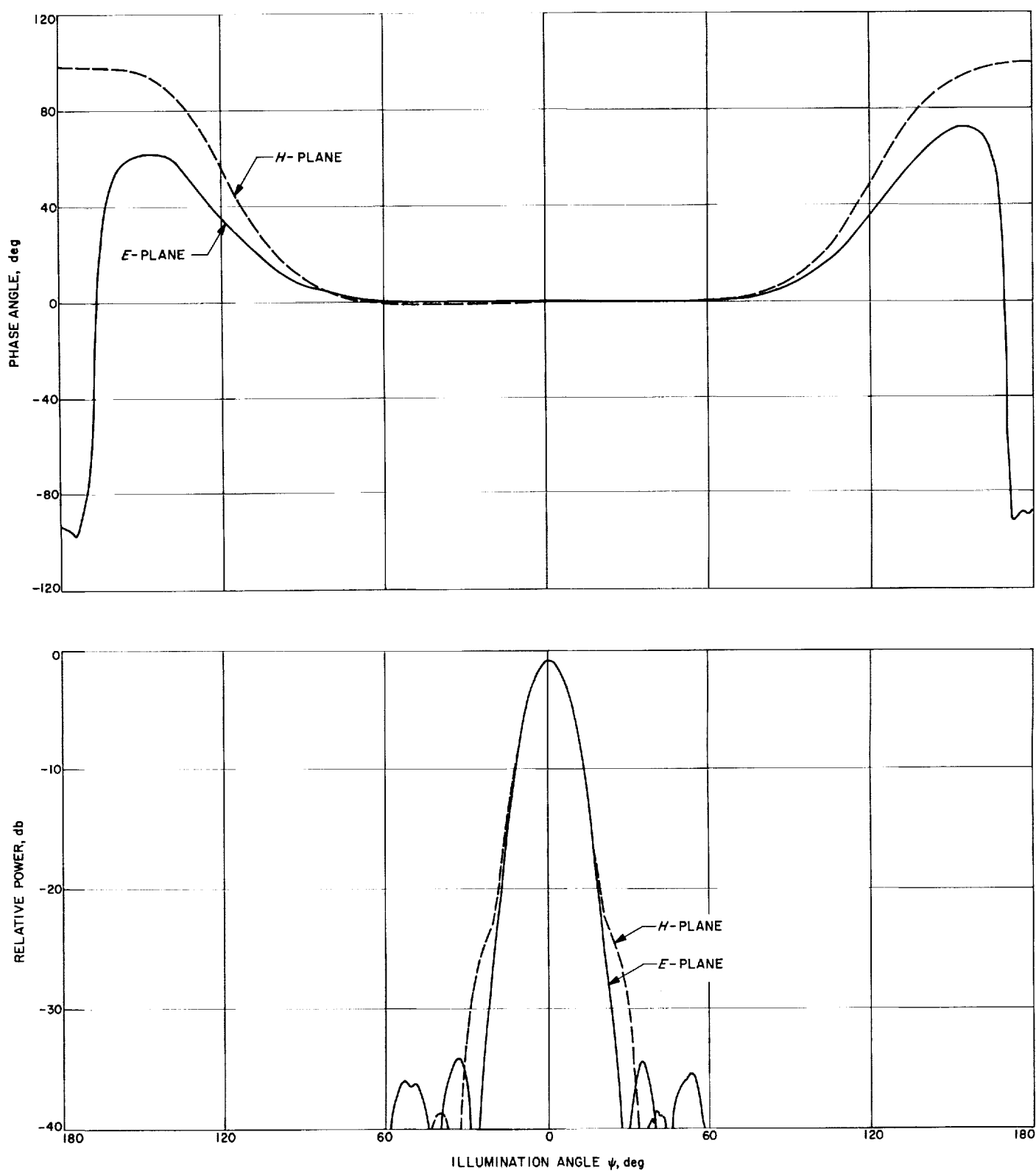
Fig. 34. Dual-mode horn, *E*- and *H*-Plane amplitude and phase patterns

RELATIVE POWER, db

PHASE ANGLE, deg

ILLUMINATION ANGLE ψ, deg

E-PLANE

H-PLANE
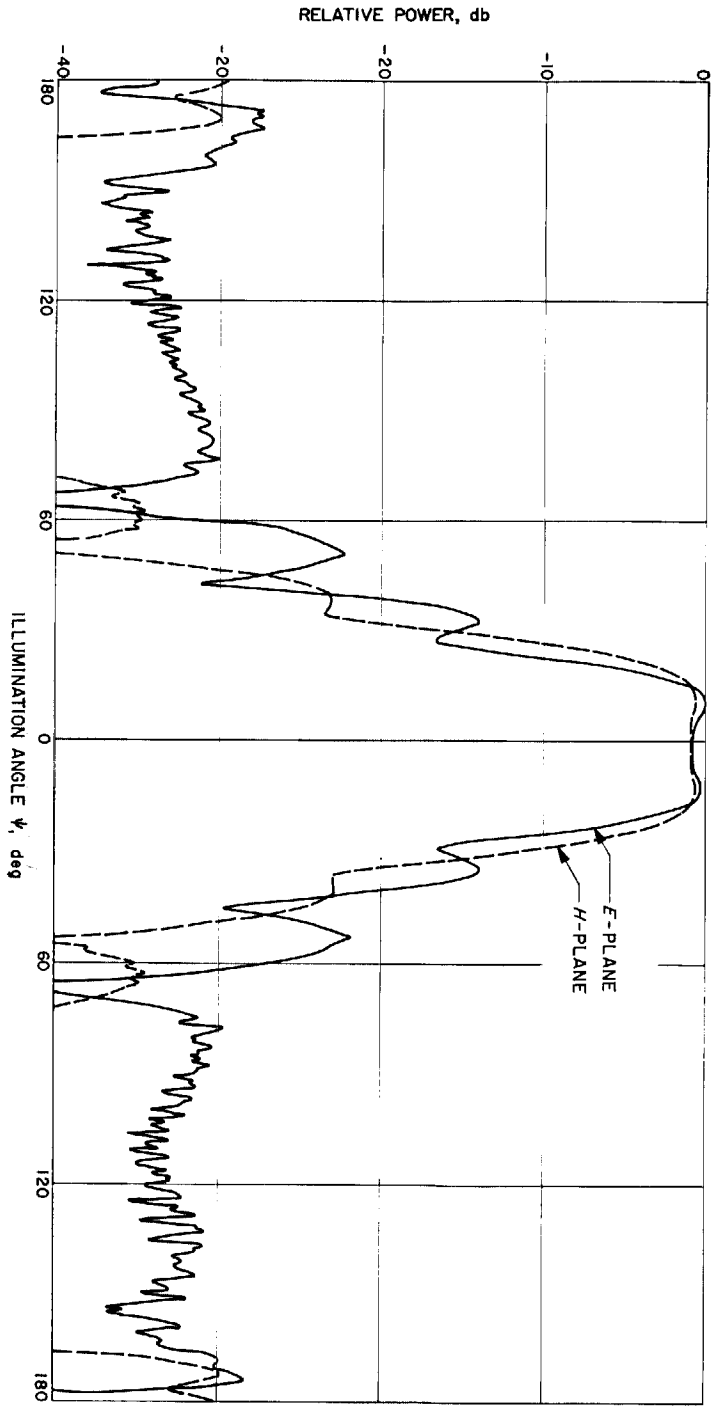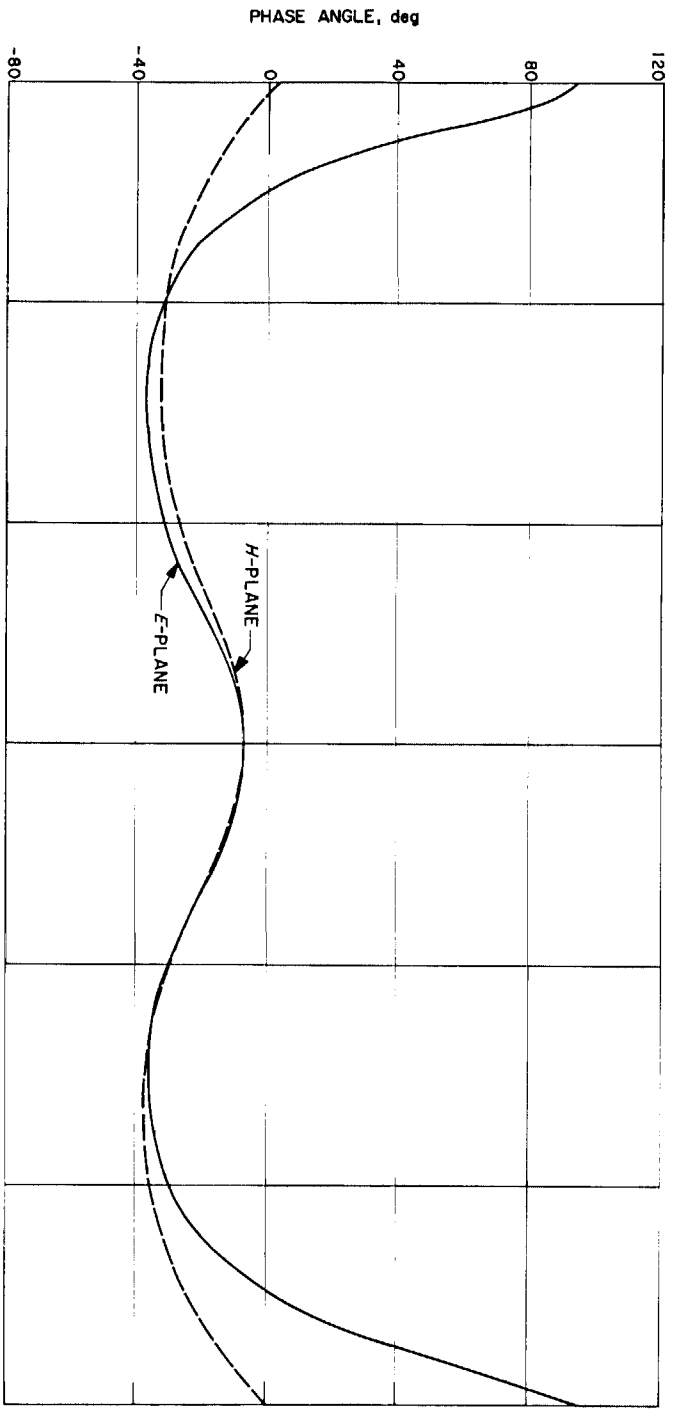
E-PLANE

H-PLANE

Fig. 35. Tri-mode horn, E- and H-Plane amplitude and phase patterns
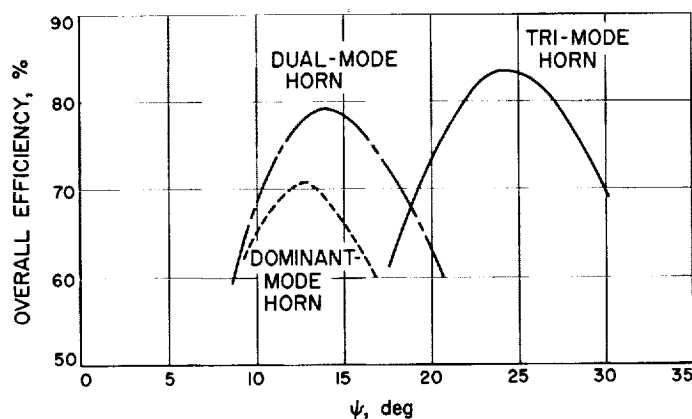
Fig. 36. Efficiency for three experimental cases

Fig. 37. Loss breakdown for three experimental cases

D. RF Techniques

T. Otoshi, C. Stelzried, and B. Seidel

1. Blue Dot Type N Connector Modifications,

T. Otoshi

a. *Summary.* The development and design details of the Blue Dot Type N connector pair were previously reported in Refs. 17 and 18. The Blue Dot Type N connector (also referred to as Blue Dot connector) is a precision modified Type N connector whose mated pair configuration, excluding the dielectric center conductor supports, approaches a discontinuity-free transmission line connection. Test results indicated that a mated pair of these connectors resulted in a VSWR of 1.02 or less at particular frequencies used at JPL (2295 Mc and 2388 Mc). It was also shown that the mating of this precision connector to precision tolerance Standard Type N connectors would result in good electrical characteristics.

This report discusses significant modifications recently made in the Blue Dot connector design. These changes may be seen in Fig. 38 and are:

(1) The 0.2760 in. OD dimension was changed to 0.2763 in. so that the characteristic impedance of Blue Dot connectors would more nearly approach the theoretical 50-ohm value.

(2) The 0.202 in. critical mating dimension was changed to 0.207 in. to make Blue Dot connectors compatible

with all Standard Type N connectors (male and female) fabricated to MIL-C-71 tolerances.

Reasons for these modifications are discussed in the following. Test data are also provided to show differences in the electrical characteristics of Blue Dot connectors made to the 0.202 in. and 0.207 in. critical mating dimensions.

b. *Modification to outer diameter dimension.* Since some of the high-precision connectors are being designed to accuracies of better than 0.1%, it becomes important whether they are dimensioned to give exactly 50 ohms in air or vacuum and at what frequency. It was pointed out that if the free-space-zero skin depth formula for Z_0 were used, then the skin depth correction cancels the dielectric constant of air correction in the frequency range where the connectors are most critically used.[1]

---

[1] R. C. Powell, Radio Standards Laboratory, National Bureau of Standards, private correspondence.

The zero-skin depth formula for the characteristic impedance of a coaxial line, Z_0, in a vacuum is:

$$Z_0 = \frac{1}{2\pi} Z_v \ln\left(\frac{b}{a}\right), \text{ ohms}$$

where

$$Z_v = \left(\frac{\mu_0}{\epsilon_0}\right)^{1/2} = 376.7303, \text{ ohms}$$

$$\mu_0 = 4\pi \times 10^{-7}, \text{ henrys/meter}$$

$$\epsilon_0 = \frac{1}{\mu_0 C^2} = 8.854\,186 \times 10^{-12}, \text{ farads/meter}$$

Then

$$Z_0 = 59.95849 \ln\left(\frac{b}{a}\right), \text{ ohms}$$

or

$$Z_0 = 138.0595 \log_{10}\left(\frac{b}{a}\right), \text{ ohms}$$

which gives for a 50.00000-ohm line in vacuum

$$\frac{b}{a} = 2.302305$$

For Blue Dot connectors:

$$a = 0.1200 \text{ in.}$$

so that

$$b = 0.2763 \text{ in.}$$

c. *Modification to critical mating dimension.* The new critical mating dimension of the Blue Dot connector may be seen in Fig. 38. The mating configuration and class tolerances for the Blue Dot connector are shown in Figs. 39 and 40 for reference purposes.

REVISIONS	
LTR	DESCRIPTION
⟨A⟩	WAS $0.2760^{+0.0005}_{-0.0003}$ D
⟨B⟩	WAS 0.202
⟨C⟩	WAS $0.120^{+0.0004}_{-0.0001}$ D

ALL DIMENSIONS IN INCHES

MATING SURFACE OUTER CONDUCTOR

STEP SURFACE OF CONTACT PIN

0.2763 ±0.0004 D ⟨A⟩

0.1200 ±0.0003 D ⟨C⟩

0.207 NOMINAL ⟨B⟩
(SEE CHART A, COLUMN (1) OF FIG. 40 FOR TOLERANCES

0.3182 ±0.0004 D

$0.3182 ^{+0.0005}_{-0.0003}$ D

0.0650 ±0.0003 D

BERYLLIUM COPPER CONTACT 0.006 SLOT × 0.330 DEEP (6 PLACES, 60 deg APART)

MALE CONNECTOR
(COUPLING NUT AND LOCK RING REMOVED FOR CLARITY)

FORWARD SURFACE OF CONTACT PIN

0.207 NOMINAL ⟨B⟩ [SEE CHART A, COLUMN (2) OF FIG. 40 FOR TOLERANCES]

MATING SURFACE OUTER CONDUCTOR

0.031 FLAT (0.3182 D-REF)

0.1200 ±0.0003 D ⟨C⟩ ⟨A⟩
0.2763 ±0.0004 D

0.0650 ±0.0003-D HOLE
0.008 SLOT × 0.125 DEEP (4 PLACES, 90 deg APART)

FEMALE CONNECTOR

Fig. 38. Blue Dot Type N connectors, critical mating dimensions; courtesy of MCC

DATUM PLANE (MATING SURFACE OUTER CONDUCTORS)

RUGGEDIZING STEP
STEP ON MALE CONTACT
MINIMIZES POSSIBLE ROCKING
UNDER SEVERE ENVIRONMENT
AND IS AN ADDED INSURANCE
FEATURE FOR PROPER ALIGNMENT

0.207 REF

0.031 FLAT REF (FEMALE CONNECTOR)
PROVIDES PROPER ALIGNMENT AND CONCENTRICITY
OF MALE CONNECTOR OUTER CONDUCTOR

A

FEMALE CONNECTOR

0.2763-D REF
0.1200-D REF

MALE CONNECTOR

VIEW A-A

A

OUTER CONDUCTOR MATING
BERYLLIUM COPPER CONTACT FINGER DESIGNED TO
MATE WITH FEMALE CONNECTOR WITH MINIMUM
DISCONTINUITY. MATING DIAMETERS ARE COMPATIBLE
DIMENSIONALLY THEREBY MAINTAINING A NOMINALLY
PERFECT COAXIAL LINE (THE CONTACT IS AN
INTEGRAL PART OF THE MALE CONNECTOR BODY)

CLOSE TOLERANCE CONDITION
BETWEEN MALE CONNECTOR
BODY AND COUPLING NUT
PROVIDES A MORE RUGGED
COUPLING AND PREVENTS
COCKING

CENTER CONDUCTOR MATING
DESIGNED TO MATE WITH A NOMINAL BUTT CONDITION. MALE PIN
AND FEMALE CONTACT HOLE DIAMETERS ARE COMPATIBLE
DIMENSIONALLY THEREBY PROVIDING A NOMINALLY PERFECT COAXIAL
LINE. SEE CHART B, FIG. 40 FOR POSSIBLE CONTACT GAP

SEE FIG. 38 FOR CRITICAL MATING DIMENSIONS
ALL DIMENSIONS IN INCHES

Fig. 39. Mating configuration of the Blue Dot Type N connector pair

ALL DIMENSIONS IN INCHES

COLOR CODE
DOT (REF)
(SEE CHART)

TOLERANCE CLASS METAL STAMPED
1/16 LETTERS (SEE CHART)

CHART A

CONNECTOR DESIGNATION	N-BD		N-STD	
COLOR CODING	BLUE		NONE	
CONNECTOR SEX	MALE (1)	FEMALE (2)	MALE (3)	FEMALE (4)
NOMINAL LOCATION	0.207	0.207	0.223	0.197
TOLERANCE CLASS A	+0.002 −0.000	+0.000 −0.002	±0.002	±0.002
TOLERANCE CLASS B	+0.003 −0.000	+0.000 −0.003	±0.005	±0.005
TOLERANCE CLASS C	+0.004 −0.000	+0.000 −0.004	±0.009	±0.010

CHART B		
CENTER CONTACT GAP AS A RESULT OF MATING SAME CLASS N-BD CONNECTORS		
CLASS	MINIMUM	NOMINAL	MAXIMUM
A	0.000	0.000	0.004
B	0.000	0.000	0.006
C	0.000	0.000	0.008

Fig. 40. Contact pin location classes for Blue Dot and Standard Type N connectors; courtesy of MCC

For reasons stated in Refs. 17 and 18, a mating dimension of 0.202 in. was previously selected for Blue Dot Type N connectors. It was experimentally shown that Blue Dot connectors made to a 0.202 in. mating dimension and mated to opposite sex Standard Type N connectors would be electrically superior to a mated pair of precision Standard Type N connectors. One disadvantage of the Blue Dot connectors with the 0.202 in. critical mating dimension was that Blue Dot male connectors would not mate with Standard Type N female connectors made to +0.010/+0.005 in. tolerances of the nominal 0.197 in. dimension specified by MIL-C-71.

During the early phase of development, the merits of a 0.207 in. dimension had been considered against the merits of a 0.202 in. design. The advantage of a 0.207 in. mating dimension is that Blue Dot Type N connectors made to this dimension will mate with all present Standard Type N connectors fabricated to MIL-C-71 tolerances. However, it was felt that Blue Dot connectors with the 0.202 in. dimension would be electrically superior to those made to the 0.207 in. dimension over a broader band of frequencies when these precision connectors were mated to the opposite sex Standard Type N connectors.

It was recently learned that a steering and working committee for the military is now in the process of preparing new performance specifications for RF connectors. This work is currently being done by the Defense Electronics Supply Center (DESC-EMM) at Dayton, Ohio, and the ASA C83.2 Task Group associated with the American Standards Association and the Electronics Industries Association. The intent of this committee is to establish two classes of connectors for each RF connector series: (1) a general class for ordinary performance, and (2) a precision class for high performance systems but not of the laboratory precision grade. The performance specifications being prepared for RF connectors will include standardization of overall envelope sizes and mating face dimensions.

It was brought to our attention that the military steering committee is seriously considering the adoption of the 0.207 in. mating dimension for future general class and precision class Type N connectors. Since Blue Dot male connectors with a 0.202 in. mating dimension will not mate with the female Type N connectors made to 0.207 in. dimension, it was decided that for future compatibility reasons the possibility of changing the Blue Dot design to accept the 0.207 in. dimension be seriously considered. The following tests were made to compare the electrical

characteristics of Blue Dot connectors made to 0.207 in. and 0.202 in. critical mating dimensions.

d. Test results. A coaxial reflectometer was used to measure the reflection coefficients resulting from various mated pairs of connectors at 2388 Mc. The methods for connector measurements using coaxial reflectometers were similar to those described in papers by W. E. Little of NBS (Refs. 19 and 20).

One of the reflectometer tests involved determination of the reflection coefficients resulting from a mating of Blue Dot and Standard Type N connectors. The mating configurations under test may be seen in Figs. 41 and 42. For these tests, the center conductor of the Blue Dot connector was not dielectrically supported. The unsupported center conductor design used with the sliding termination assembly permitted the critical center contact pin location to be varied and, therefore, facilitated changing the critical mating dimension to desired values. As may be seen from Figs. 41 and 42, a change in the critical mating dimension will result in a change in the center contact gap. Figs. 43 and 44 show graphs of the measured mated pair VSWR versus center contact gap for the two mating configurations tested. For a mated condition of female Standard Type N and male Blue Dot Type N connectors, the center contact gap spacings for the 0.202 in. and 0.207 in. cases are 5 and 10 mils, respectively. The measured VSWR's for the respective cases were 1.035 and 1.040 (see Fig. 43). For the mated pair of male Standard Type N and female Blue Dot Type N connectors, the resulting center contact gaps are 21 and 16 mils for the 0.202 in. and 0.207 in. cases, respectively. It may be seen from Fig. 44 that the resulting VSWR's were 1.048 for both the 0.202 in. and 0.207 in. cases. The accuracy of the measured VSWR values is estimated to be ±0.7%.

A VSWR of 1.018 was measured at 2388 Mc for a mated pair of 0.207 in. Blue Dot Type N connectors and was the same VSWR value as that for a mated pair of 0.202 in. Blue Dot Type N connectors (within the estimated accuracy of this measurement of ±0.005). For these tests, Blue Dot adapters with 0.390 in. diameter Teflon bead supports described in Ref. 17 were used. Fig. 45 shows the test configuration and reflectometer method used to measure the reflection coefficient of a mated pair of connectors having center conductor supports. A necessary assumption which was made for this latter measurement was that the effects of the reflection coefficient due to a connection of beadless Blue Dot connectors produced negligible error in the measurement (see Fig. 45).
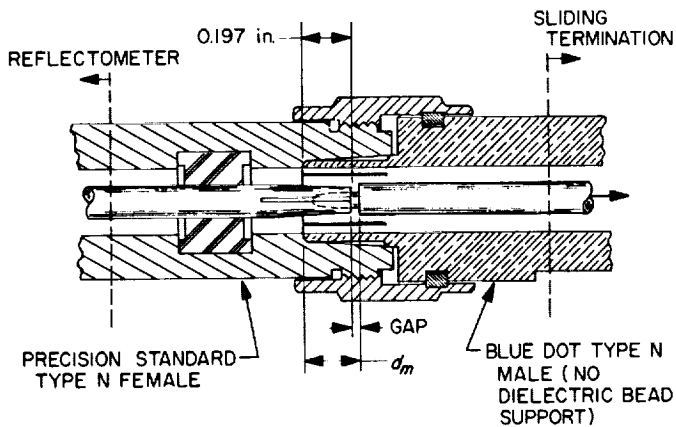
NOTES:
1. d_m = CRITICAL MATING DIMENSION FOR THE MALE BLUE DOT CONNECTOR AND MAY BE VARIED

2. GAP = d_m − 0.197 in.

Fig. 41. Mating test configuration of Standard Type N female and Blue Dot Type N male connector
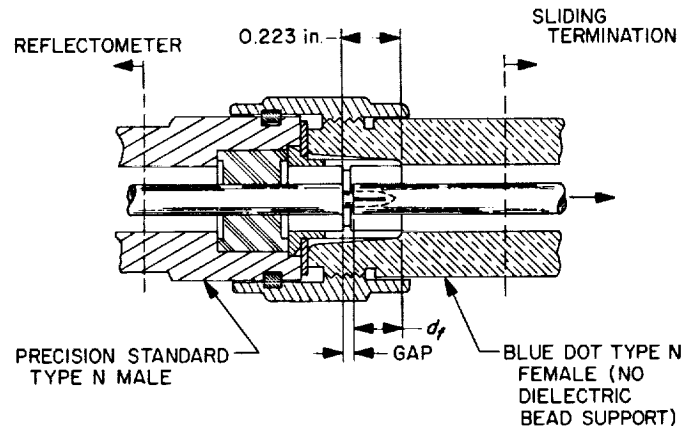
NOTES:
1. d_f = CRITICAL MATING DIMENSION FOR THE FEMALE BLUE DOT CONNECTOR AND MAY BE VARIED

2. GAP = 0.223 in. − d_f

Fig. 42. Mating test configuration of Standard Type N male and Blue Dot Type N female connectors

Fig. 43. VSWR versus center contact gap at 2388 Mc for a mating of a Standard female Type N connector and a Blue Dot male Type N connector; no dielectric bead support was used for the Blue Dot connector

Fig. 44. VSWR versus center contact gap at 2388 Mc for a mating of a Standard male Type N connector and a Blue Dot female Type N connector; no dielectric bead support was used for the Blue Dot connector

SLIDING TERMINATION
ELEMENT

FEMALE BLUE DOT

MALE BLUE DOT

TUNER T_x

(a) REFLECTOMETER TUNING

CRITICAL MATING
DIMENSION

SLIDING TERMINATION
ELEMENT

TUNER
T_x

BLUE DOT MATED PAIR # 1

$\left[\Gamma_1 \text{ ASSUMED TO BE}\right.$
IDENTICAL TO THAT IN
STEP (a) AND IS TUNED
$\left. \text{OUT}\right]$

BLUE DOT MATED PAIR # 2

(CONNECTOR PAIR UNDER
TEST)

BLUE DOT MATED PAIR # 3

$\left[\Gamma_3 \text{ ASSUMED TO BE}\right.$
IDENTICAL TO THAT IN
STEP (a) AND THAT

$\left. |\Gamma_3| \ll |\Gamma_2| \right]$

(b) MEASUREMENT

Fig. 45. Measurement method for testing Blue Dot Type N connectors with supported center connectors

The advantage of the 0.207 in. design in offering unrestricted compatibility to all present Standard Type N connectors (per MIL-C-71) were considered against the slightly better electrical characteristics of the 0.202 in. design at 2388 Mc. It was decided that future Blue Dot connectors will be fabricated to have the critical mating dimension of 0.207 in.

2. DC Potentiometer Insertion Loss Test Set,
C. Stelzried

a. Summary. The dc comparator microwave insertion loss test set described in Ref. 21 has been adapted for use at 90 Gc. This equipment is being used in the calibration of the 90 Gc radiometer described in Ref. 22.

b. Recent work. The Hewlett Packard Company has extended the frequency range of the Model 431B power meter to 40 Gc with the introduction of the thermister mount model R 486A. These mounts are being used in an insertion loss test set at 90 Gc by using commercial waveguide transitions from RG 96/U (thermister mount)

to the millimeter wave RG 99/U (Fig. 46). The VSWR looking toward A and A' is less than 1.04 (cf Fig. 46). The thermistor mounts are not a very good match because they are operated outside of their rated frequency range; a 3-db attenuator is used before the thermistor mount input to reduce the VSWR at point A'. The VSWR at A is not critical with the match of the thermistor mount due to the directivity of the 10 db coupler. The signal source used in the test set is a Raytheon Model 673 klystron. A photograph of the waveguide components is shown in Fig. 47. Although this configuration is not as stable as it is at X- and S-band frequencies, it is quite satisfactory. Fig. 48 shows a recording of the stability measured at the recorder output of the null detector. The principal source of the instability has not been determined with certainty, it is most likely due to frequency instabilities of the klystron. The calibration on Fig. 48 was obtained by momentarily adjusting the precision divider by an amount corresponding to 0.1 db change in insertion loss. The resolution obtained to date is about 0.02 db, which will also be the accuracy for small values of insertion loss. The linearity has not been evaluated although it should remain the same as measured at S-band, better than 0.1% to 6 db.

Fig. 46. Block diagram of 90-Gc insertion loss test set

Fig. 47. Photograph of 90-Gc dc potentiometer insertion loss test set

3. Simultaneous Lobing Radiometric Tracking System, *B. Seidel*

a. Summary. The S-band systems for the DSIF 85-ft and future 210-ft antennas will use simultaneous lobing angle tracking feed systems. A radiometer which could be used with the tracking feed would be a useful device for angle pointing and gain calibration of the antenna system using radia star sources. An X-band laboratory model of a simultaneous lobing receiver channel has been constructed. A noise tube is used to simulate the signal from a radio source. Preliminary experiments have been performed to examine the angle detector output versus reference/error channel differential phase shift and error channel noise-to-signal ratio.

b. Recent work. A simultaneous lobing radiometric system would allow one to automatically track a celestial noise source. The a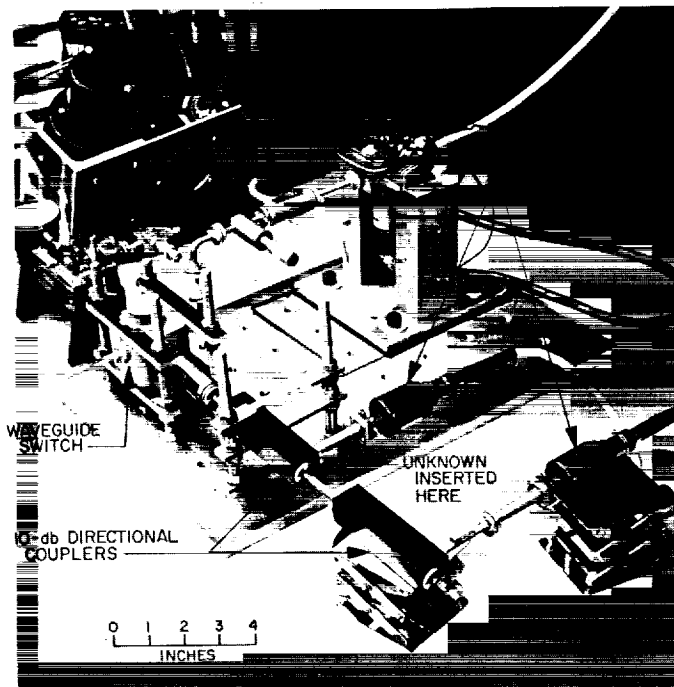pparent source position as seen by the antenna could then be compared with the known source position, and the RF pointing error directly determined; also, with the tracking capability, a sum channel gain measurement using the radio source could be operationally simplified.

Though an operating system would operate at the DSIF S-band frequency, X-band was chosen for a laboratory mockup because of the relative ease of handling the
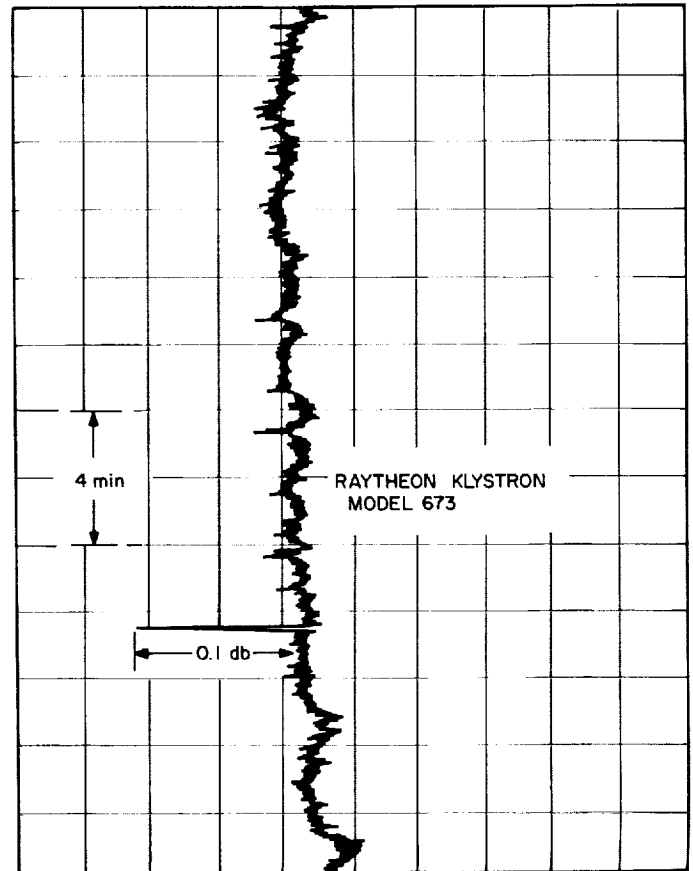
Fig. 48. 90-Gc insertion loss test set stability recording

components and the availability at JPL of these components. A block diagram of the experimental simultaneous lobing receiver channel is shown in Fig. 49. All of the components in the system are commercially available with the exception of the broadband 30 Mc phase detector. The phase detector is a modification of the 10 Mc unit in use at the CIT Owens Valley Radio Observatory; the design drawings for it were supplied by Dr. R. Read of CIT.

Fig. 50 shows the normalized output of the phase detector for various noise-to-signal (N/S) ratios on the inputs of the sum and difference arms and a constant 3-db difference in N/S ratio between them; the signal is white noise obtained from a HP X347A noise source. Fig. 51 is a similar family of curves, but with a constant 10-db difference in N/S ratio between the two inputs. The various N/S ratios are obtained by adjusting the variable attenuators at the input and in the difference arm of the receiver. The shifts in the dc level of the various curves with input N/S can probably be reduced by selection of detector diodes.
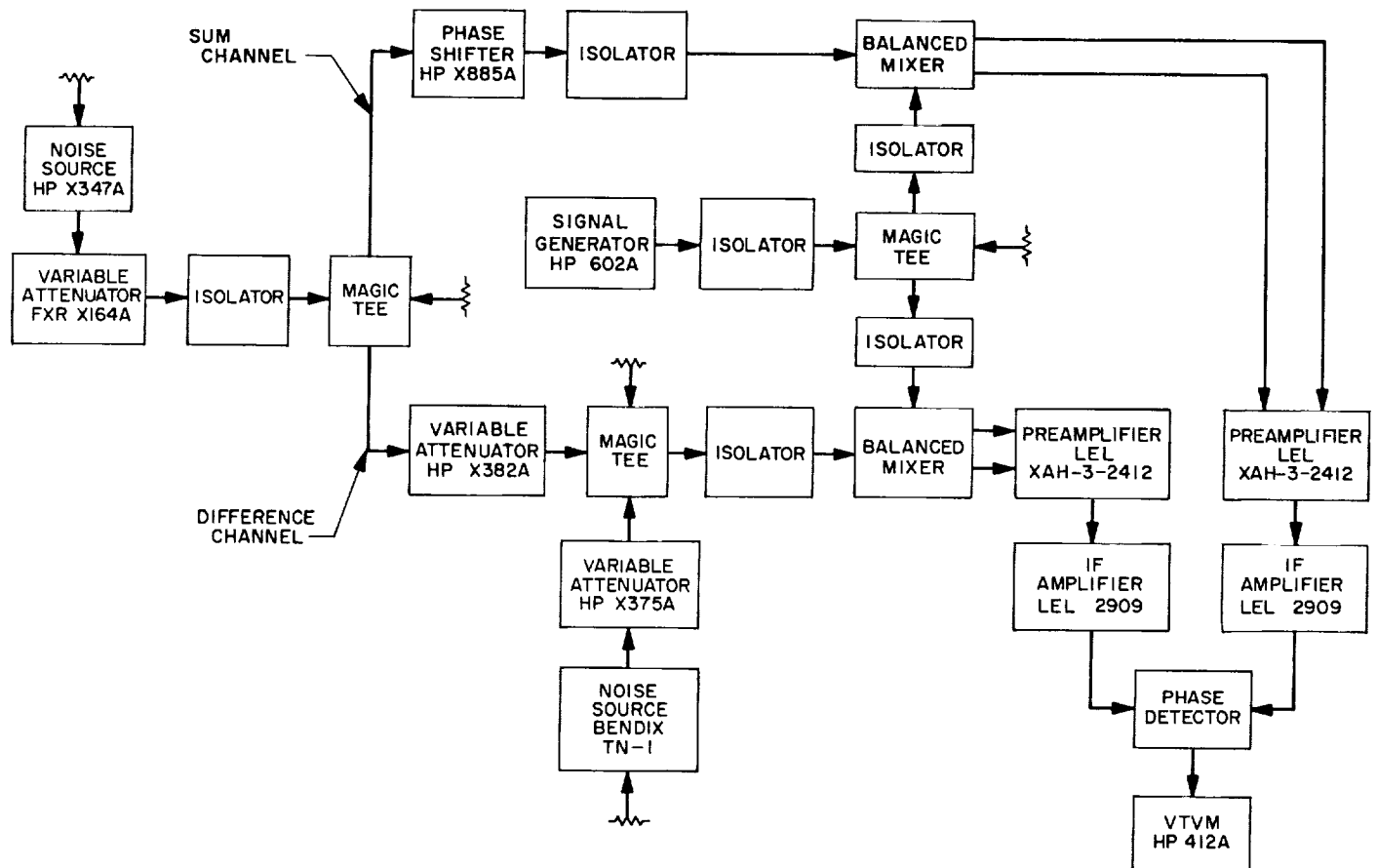
Fig. 49. Block diagram of laboratory mockup of monopulse tracking station

SYMBOL	$\frac{N}{S}$ RATIO Σ ARM, db	$\frac{N}{S}$ RATIO Δ ARM, db	NORMALIZATION FACTOR
◇	-3 db	0 db	6.50
□	0 db	3 db	4.60
△	3 db	6 db	2.125
▽	6 db	9 db	1.50
○	9 db	12 db	0.62
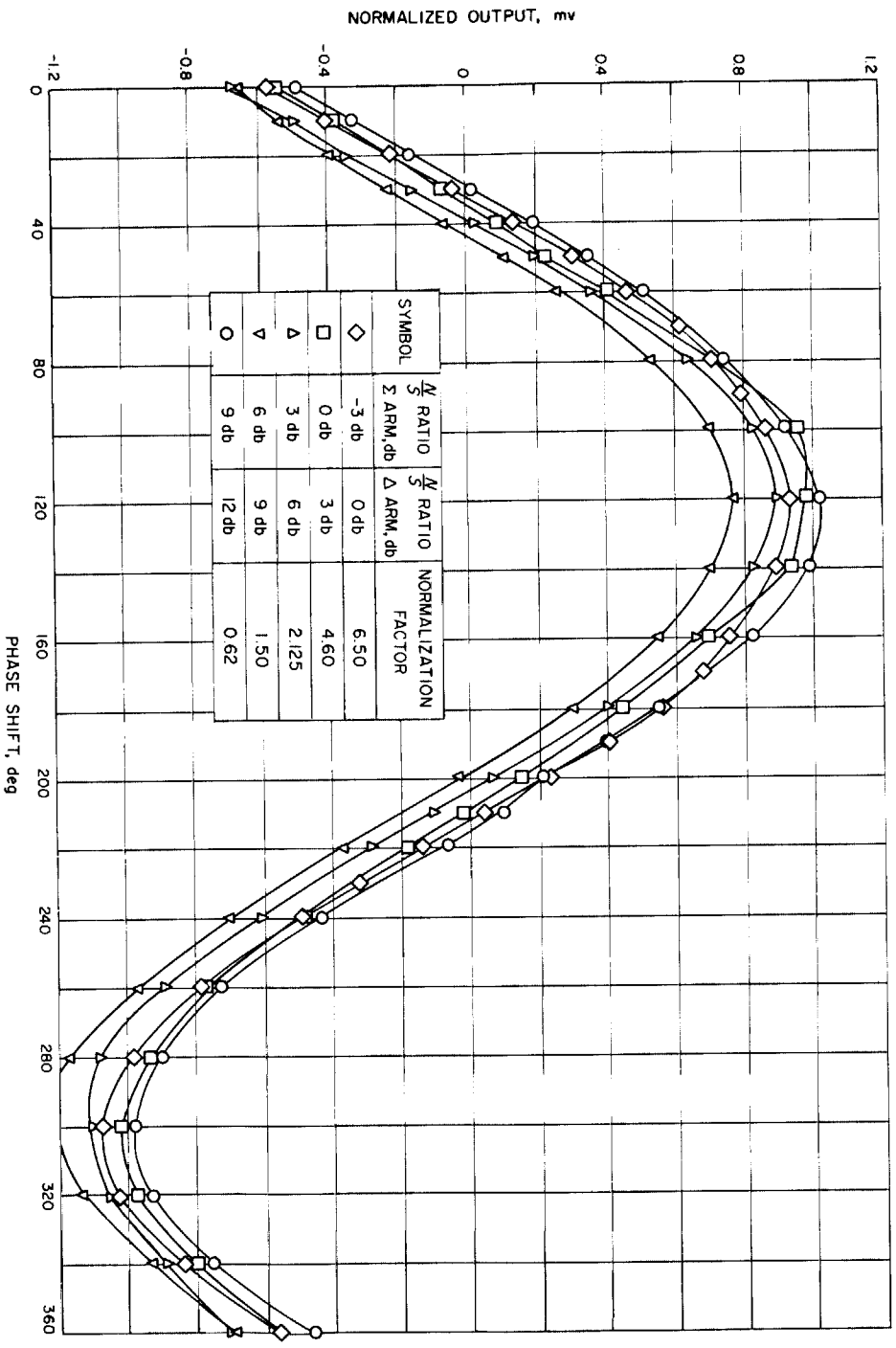
Fig. 50. Normalized detector output versus phase shifter setting for 3-db N/S difference between Σ and Δ arms
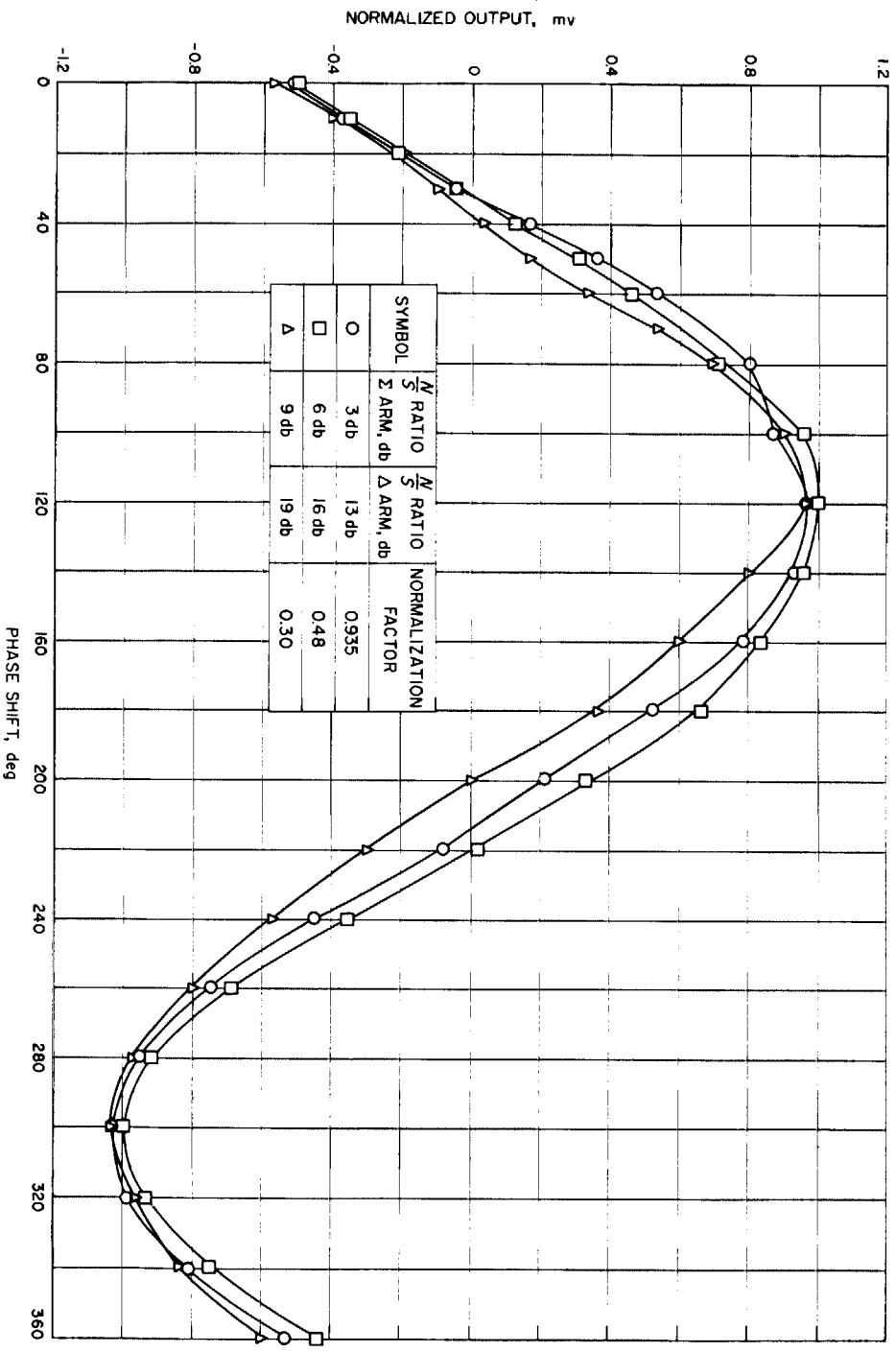
Fig. 51. Normalized detector output versus phase shifter setting for 10-db N/S difference between Σ and Δ arms

The table within the figure reads:

SYMBOL	$\frac{N}{S}$ RATIO Σ ARM, db	$\frac{N}{S}$ RATIO Δ ARM, db	NORMALIZATION FACTOR
O	3 db	13 db	0.935
□	6 db	16 db	0.48
Δ	9 db	19 db	0.30

Fig. 52 shows detector output versus N/S ratio in the difference channel of the receiver with the sum channel held at a constant N/S ratio and the phase shifter set for maximum output.

Table 5 shows the bias error developed by the detector as a function of input signal level. This error can be balanced out when operating at a nearly constant input signal level, which would be the normal situation in tracking a radio source. The question then is the detector bias stability over the length of time of a measurement. This characteristic will be evaluated.
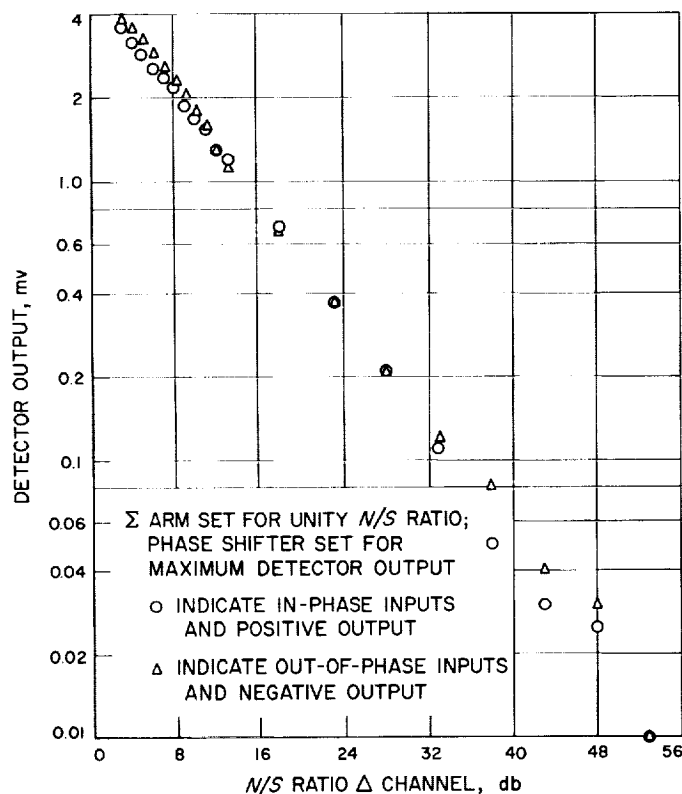
Table 5. Detector bias error

FXR, db	Output, mv	N/S: — Σ, db	N/S: — Δ, db
0	0.0	—3	7
1	—0.05	—2	8
2	—0.10	—1	9
3	—0.13	0	10
4	—0.16	1	11
5	—0.18	2	12
6	—0.19	3	13
7	—0.20	4	14
8	—0.21	5	15
9	—0.21	6	16
10	—0.22	7	17
11	—0.22	8	18
12	—0.22	9	19
13	—0.22	10	20
14	—0.22	11	21
15	—0.22	12	22
16	—0.23	13	23
17	—0.23	14	24
18	—0.23	15	25
19	—0.23	16	26
20	—0.23	17	27
25	—0.23	22	32
30	—0.23	27	37
35	—0.23	32	42
40	—0.23	37	47
50	—0.23	47	57

Each channel noise temperature 2400°K; phase shifter at 32°; phase detector nulled.

Fig. 52. Detector output versus N/S ratio on Δ arm

References

1. Stelzried, C. T., and Rusch, W. V. T., "Millimeter-Wave Work," SPS 37-25, Vol. IV, pp. 113–117, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

2. Lynn, V. L., Meeks, M. L., and Sohigian, M. D., Radiometric Measurements at 8.5-mm Wavelength with a 28-foot Antenna During December 1962, Technical Report 330, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., October 8, 1963.

3. Stelzried, C. T., "A Liquid-Helium Cooled Coaxial Termination," Correspondence, Proceedings of the IRE, Vol. 49, No. 7, July 1961.

References (Cont'd)

4. Stelzried, C. T., "Temperature Calibration of Microwave Terminations," *SPS 37-25*, Vol. IV, pp. 118–121, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

5. Ramo, S., and Whinnery, J., *Fields and Waves in Modern Radio*, 2nd Ed, Table 9.01, John Wiley & Sons, 1953.

6. Siddoway, J., and Wells, W. H., "Optical and Infrared Masers," *SPS 37-24*, Vol. IV, pp. 137–148, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.

7. Erpenbach, H., and Finnie, C., "Superconducting Bolometers," *SPS 37-24*, Vol. IV, pp. 148–150, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.

8. Potter, P. D., "Antenna Feed Research," *SPS 37-24*, Vol. IV, pp. 150–154, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.

9. Stratton, J. A., *Electromagnetic Theory*, p. 405, McGraw-Hill Book Company, New York, 1941.

10. Silver, S., "Microwave Antenna Theory and Design," *MIT Radiation Laboratory Series, Vol. 12*, McGraw-Hill Book Co., New York, 1949.

11. Bathker, D., "Radiation Pattern and Antenna Efficiency Studies," *SPS 37-23*, Vol. III, pp. 36–38, Jet Propulsion Laboratory, Pasadena, California, September 1963.

12. Potter, P. D., "A New Horn Antenna With Suppressed Sidelobes and Equal Beamwidths," *The Microwave Journal*, June 1963

13. Ludwig, A. C., "Multimode Horn Research," *SPS 37-24*, Vol. IV, pp. 154–156, Jet Propulsion Laboratory, Pasadena, California, December 1963.

14. Rusch, W. V. T., "Scattering from a Hyperboloidal Reflector in a Cassegrainian Feed System," *IEEE Transactions on Antennas and Propagation*, Vol. AP-11, No. 4, July 1963

15. Potter, P. D., "Spherical Wave Functions," *SPS 37-24*, Vol. IV, pp. 150–154, Jet Propulsion Laboratory, Pasadena, California, December 1963.

16. Ludwig, A. C., "Multimode Feedhorns," *SPS 37-19*, Vol. IV, pp. 190–196, Jet Propulsion Laboratory, Pasadena, California, February 1963.

17. Otoshi, T., "Development of the Blue Dot Type N Connector," *SPS 37-21*, Vol. IV, pp. 194–202, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.

18. Otoshi, T., "Blue Dot Type N Connectors," *SPS 37-22*, Vol. IV, pp. 199–203, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.

19. Little, W. E., "A Coaxial Reflectometer System," *Meeting on High-Precision Connectors*, Report 7277, pp. 14–28, National Bureau of Standards, Boulder, Colorado, June 29, 1961.

20. Little, W. E., "Measurement of the Insertion VSWR of Coaxial Connectors Using Reflectometer Techniques," *Meeting on High-Precision Connectors*, Report 7277, pp. 29–40, National Bureau of Standards, Boulder, Colorado, June 29, 1961.

References (Cont'd)

21. Petty, S., and Stelzried, C. T., "Dc Potentiometer Radio Technique," *SPS 37-22*, Vol. IV, pp. 196–198, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.

22. Stelzried, C. T., "Dc Potentiometer Insertion Loss Test Set," *SPS 37-25*, Vol. IV, pp. 132–133, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

XVIII. Communications Systems Research

A. Coding Theory

E. C. Posner, G. Solomon, and R. E. Block

1. A Code Over the 32 Element Field for Use on Interstation Teletype Channels,

E. C. Posner and G. Solomon

a. Abstract. Commands to be sent to Goldstone and other DSIF Stations for transmission to spacecraft are computed at JPL and sent over teletype channels. It is essential that the probability of error be kept extremely small. This article describes a coding scheme for use on these channels, which combines simplicity and speed with extremely low error probability. In fact, output bit-error probability of less than 10^{-8} is obtained with a rate 0.274 system requiring no re-transmissions, when the input binary symbol error probability is 0.02.

b. Introduction. The teletype channel format consists of 63 columns on a punched tape. Each column contains 5 positions which may be "punched" or "not punched."

Thus, the natural word length is 315. Of these bits, 5 are used for synchronization information, so that the data word is of length 310. We shall therefore consider coding schemes which have word lengths of about 310. (One would be willing to lose a few check bits, not transmitting them, if it results in greater coding and decoding simplicity.) In addition to the constraints that the coder and decoder be reasonably easy to build and to operate, and that the decoding be performable in real time, it is also desirable that the coder and decoder be essentially the same machine. This will allow the overseas stations to communicate with JPL using the same coding system, without the necessity for having two separate devices at each station.

The aim is for a binary decoded symbol-error probability of $\leq 10^{-8}$. Recent measurements indicate that the worst-case uncoded symbol-error probability could be as high as 2×10^{-2} on the overland portion. It is also shown that the errors occur independently, and that the transition $0 \rightarrow 1$ is as like as $1 \rightarrow 0$. We assume throughout the remainder that the teletype channel is a binary

symmetric channel, with an input symbol-error probability $p = 0.02$, which is considered to be a *large* error probability.

In the subsequent parts of this article, we describe a coding system that can be used, and indicate the information rates and error probabilities associated with the different information rates possible. To calculate the output *information-bit* error probability of each system, the output *word*-error probabilities were calculated, and a technique described in Ref. 1 to translate these figures into output information-bit error probabilities was used.

Since the overseas links can exhibit burst-error characteristics, the burst-error detecting capabilities of the codes will also be indicated. The possible modes of operation of the code will also be indicated. But one may wish to use the codes as error-detecting codes, with re-transmissions allowed. But to avoid having to ask for re-transmissions too frequently when errors are detected, one could decide to correct all single errors, and detect all errors that can still be detected, as in Ref. 2. This choice can cut the re-transmission rate quite substantially, and perhaps still achieve the required output error rate. (However, this is not the case in the present channel, as will be seen.) The decoder would be switched to perform in one of several modes as the channel characteristics change. A particular class of codes will now be discussed.

c. Codes over the 32 element field. The use of these codes is the main innovation of this paper. Let us consider cyclic codes of length 63 over the field of $32 = 2^5$ elements, $GF(2^5)$. A $(63, k)$ cyclic code, for example, is generated by a linear difference equation of degree k with coefficients in $GF(2^5)$. That is, we have

$$a_{k+n} + \alpha_1 a_{k+n-1} + \alpha_2 a_{k+n-2} + \cdots \alpha_n a_k = 0,$$

where $\alpha_i \in GF(2^5)$, $i = 1, 2, \cdots, n$; $n = 0, 1, 2, \cdots$, and $a_0, a_1, a_2, \cdots, a_{k-1}$ are given elements of $GF(2^5)$. For the code to be cyclic and of length 63, the associated polynomial

$$f(x) = x^k + \alpha_1 x^{k-1} + \alpha_2 x^{k-2} + \cdots \alpha_{k-1} x + \alpha_k$$

must divide $x^{63} + 1$.

Now let $g(x)$ be a polynomial which divides $x^{63} + 1$, with coefficients in $GF(2)$. Let g be irreducible over $GF(2)$. It can be shown that $g(x)$ is irreducible over $GF(2^5)$. Thus, a $(63, k)$ binary cyclic code is contained in the $(63, k)$ 2^5-ary cyclic code and both are generated

from the *same* difference equation. In particular, represent any element α of $GF(2^5)$ as binary 5-tuples; e.g.,

$$\alpha \rightarrow (c_0, c_1, c_2, c_3, c_4) = c_0 + c_1 \delta + c_2 \delta^2 + c_3 \delta^3 + c_4 \delta^4,$$
$$c_i \in GF(2),$$

where δ is a primitive root of $GF(2^5)$ over $GF(2)$. We then have a very simple encoding and decoding procedure inherited from the $(63, k)$ *binary* cyclic code. Thus, the $(63, k)$ 2^5-ary code is now (via the binary representation of α) a $(315, 5k)$ *binary* code. In addition, if we shorten the $(63, k)$ cyclic code by one symbol, we obtain a $(310, 5(k-1))$ binary code with an inner structure which allows either symbol error correction or the correction of certain bursts of length 2 to 5. If one ignores the $GF(2^5)$ structure and merely considers these codes as five copies of a binary code of length 62, then one could decode the five codes in *series* with the same decoder. But not only does the $GF(2^5)$ interpretation allow in effect *parallel* decoding, but also the way in which the five codes are related allows the symbols of length 5, the teletype characters, to be detected and corrected as a unit.

Table 1 gives the values of k one obtains for these $(63, k)$ codes over $GF(2^5)$. (The code finally used is, however, $(62, k-1)$). Also listed are the recursion polynomials for generating these codes; as mentioned, the coefficients are in $GF(2)$. The residual burst-detection capability over $GF(2^5)$ when option A or B below is used is also indicated. For option A, it is just $n - k$; for option B, Ref. 2 is used.

Table 2 lists the output binary information-bit error probabilities obtained with these codes when p, the input

Table 1. Recursions for codes of length 63 over GF (2^5)

k	Generator of code (in octal)	Burst detection $n-k$, option A	Residual burst detection (Ref. 2), option B
7*	3 0 5	56	52
10*	2 7 7 1	53	50
16*	3 5 4 2 1 7	47	43
18*	1 2 0 5 6 5 5	45	40
24	1 0 2 2 6 1 7 2 3	39	34
30	1 5 6 0 4 6 7 6 2 4 7	33	28
36	1 7 2 4 7 2 1 5 0 1 5 4 1	27	21
39	1 5 3 3 2 3 7 7 7 1 7 2 5 3	24	21
45	1 4 7 5 2 2 6 3 7 1 1 4 0 4 6 3	18	13
51	1 1 7 1 5 0 6 1 5 7 6 2 2 3 5 3 0 5	12	8
57	1 7 6 5 3 1 5 6 6 4 4 7 0 5 7 1 2 1 4 1	6	1

Table 2. Output bit-error probabilities of (62, k — 1) codes over GF (2⁵)

k	Rate	Minimum distribution d	Error correction e	Re-transmission rate option A, %	Re-transmission rate option B, %	Adjusted rate option A	Adjusted rate option B	Error probability option A	Error probability option B	Error probability option C
7*	0.097	31	15	71.4	71.0	0.028	0.028	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
10*	0.145	27	13	71.4	71.0	0.041	0.042	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
16*	0.242	23	11	71.4	71.0	0.069	0.070	$< 10^{-10}$	$< 10^{-10}$	2.2×10^{-10}
18*	0.274	21	10	71.4	71.0	0.077	0.079	$< 10^{-10}$	$< 10^{-10}$	2.3×10^{-9}
24	0.371	15	7	71.4	71.0	0.106	0.108	$< 10^{-10}$	$< 10^{-10}$	1.4×10^{-6}
30	0.468	13	6	71.4	71.0	0.139	0.142	$< 10^{-10}$	1.2×10^{-10}	8.6×10^{-6}
36	0.565	11	5	71.4	71.0	0.161	0.164	1.0×10^{-10}	1.2×10^{-8}	4.2×10^{-5}
39	0.613	9	4	71.4	71.0	0.175	0.178	1.5×10^{-8}	8.5×10^{-7}	2.1×10^{-4}
45	0.710	7	3	71.4	71.0	0.203	0.206	7.4×10^{-7}	3.4×10^{-5}	7.9×10^{-4}
51	0.807	5	2	71.4	70.9	0.231	0.234	3.0×10^{-5}	7.0×10^{-4}	2.5×10^{-3}
57	0.904	3	1	71.3	70.4	0.259	0.267	6.3×10^{-4}	6.2×10^{-3}	6.2×10^{-3}

symbol-error probability for the *binary* symbols, is 0.02. The calculations were done by R. McEleice using the convenient tables of the binomial distribution in Ref. 3. The decoding options considered are defined as follows:

(A) Error detection only.

(B) Single-error correction and residual burst-error detection.

(C) Error correction up to full capability.

The definition of output error probability, as in Ref. 1 used above, is to divide the output word-error probability by the number of information bits.

Under options A and B, the probability of detected but uncorrected error must be considered, for that probability is the re-transmission frequency. It is reasonable to demand that this frequency be less than 20%. In Table 1, it is 70%, which is far too high. Thus, only option C could be used. Whenever the output bit-error probability in option C is $< 10^{-8}$, an asterisk is placed next to the value of k in Tables 1 and 2.

d. Conclusions. The code that has been adopted for use with the DSIF command teletype links is the code obtained from the (63,18) code over GF (2⁵) by deleting one information symbol. As a binary code, it is a (310, 85) code, with information rate $85/310 \approx 0.274$. Option C, full error correction out to the error-correcting capability of 10 [over GF (2⁵)] is used. The output bit-error probability is about 2.3×10^{-9}. Although higher information rates are attainable with acceptable output bit-error probability, as Table 2 shows, these options involve re-transmission rates of 70%. In fact, the rate of a code used with re-transmissions should be multiplied by one minus the re-transmission rate to get the *true* rate.

Table 2 shows this too. Then the rate of the $k = 18$ code is *higher* than the rate of any other code in the table using options A or B.

Encoding and decoding would be done as in Ref. 4, Chap. 9. The first information symbol would be zero when encoding, and the zero would not be transmitted. In decoding, the missing first symbol 0 would be put in before using the decoding procedure. Ways of doing these calculations in GF (2⁵) directly have already appeared (Ref. 5). These encoders and decoders could be built out of digital modules, or, alternatively, a general purpose digital computer could be used. The transmission rate of 6 five-symbol characters per second is slow enough to allow all computations to be done in real time.

The recursion polynomial for the (63,18) code is obtained from Table 2 as

$$x^{18} + x^{16} + x^{11} + x^9 + x^8 + x^7 + x^5 + x^3 + x^2 + 1.$$

2. Transitive Collineation Groups on Constant-Distance Codes, R. E. Block

a. Introduction. In previous articles (Block, Ref. 6; Titsworth, Ref. 7), it was shown that a binary constant-distance code closed under cyclic shifts consists of repeated cyclic shifts of some fixed code word, together, possibly, with a trivial (i.e., constant) word. In this article, this result will be shown to remain valid when the cyclic shifts are replaced by an arbitrary transitive nilpotent group G of permutations of the places.

A permutation of the places, or of the columns of a matrix whose rows are the code words, is called a *collineation* of the code or of the matrix (Block, Ref. 8).

It will be shown that the above result need not hold if G is not assumed nilpotent; this will be done by giving an infinite class of Hadamard matrices with doubly transitive collineation groups.

b. Codes with a nilpotent transitive collineation group. In this article, it is assumed that M is an $m \times n$ matrix of $+1$'s and -1's whose rows constitute the words of a constant-distance code with Hamming distance d $(d > 0)$; it is assumed that G is a group of collineations of M which is transitive on the columns. It follows from Theorem 3 of Block (Ref. 8) that G regarded as a group of permutations of the rows of M, has at most *two* orbits. If T_i is one of these orbits, let v_i, r_i and k_i denote, respectively, the number of words in T_i, the number of $+1$'s in each of the words, and the number of words in T_i with $+1$ in any column. Then if G has two orbits T_1 and T_2 on the rows, one has (Block, Ref. 9):

$$\frac{n(r_1 + r_2 - d)}{2} = r_1 r_2, \qquad (1)$$

$$r_i(k_i - 1) = \left(r_i - \frac{d}{2}\right)(v_i - 1), \qquad i = 1, 2. \qquad (2)$$

Also let H denote the subgroup of G fixing the first column. The following theorem is now obtained.

Theorem 1. Suppose that T_1 and T_2 are distinct orbits of G on the rows. For $i = 1, 2$, take α_i in T_i and let S_i be the subgroup of G fixing α_i. Suppose that p is any prime such that the highest power p^j of p dividing n does not divide d. Then, either for $i = 1$ or 2, S_i contains the normalizer of a Sylow p-subgroup of G, and $p|(v_i - 1)$, $p^j|r_i$.

Proof. If the orbit T_i is trivial (consists of a constant word) then $S_i = G$ and the conclusion is obvious. Thus, suppose that both orbits are nontrivial. Take a prime p such that p^j, the highest power of p dividing n, does not divide d. Let p^e and p^f be the highest powers of p dividing r_1 and r_2, respectively; by choice of notation one may suppose that $e \leq f$. By Eq. (1), $p^j|r_1 r_2$.

Suppose first that $p \nmid (v_1 - 1)$ and $p \nmid (v_2 - 1)$. Then by Eq. (2), $p^e|[r_1 - (d/2)]$ and $p^f|[r_2 - (d/2)]$, so that $p^f|(d/2)$ and $p^e|(r_1 + r_2 - d)$. If $p > 2$ then p^{j+e} divides the left side of Eq. (1) while p^{e+f} is the highest power of p dividing the right side. Hence, $f \geq j$, so that $p^j|d$, a contradiction. If $p = 2$ then $p^{e-1}|[(r_1 + r_2 - d)/2]$ and

p^{j+e-1} divides the left side of Eq. (1). Then $f \geq j - 1$, $p^{j-1}|(d/2)$ and $p^j|d$, again a contradiction.

Hence, $p|(v_i - 1)$ for some i, with $i = 1$ or 2. Then, since $p|([G:S_i] - 1)$, $p \nmid [G:S_i]$, and S_i contains a Sylow p-subgroup of G. Suppose that K is any subgroup of G, and consider the orbits of K when K is regarded as a permutation group on the columns. For each of these orbits there is an x in G such that the number of elements in the orbit is $[K:K \cap xHx^{-1}]$. If p^l is the highest power of p dividing $|H|$ then p^{j+l} is the highest power of p dividing $|G|$. Hence, if K contains a Sylow p-subgroup of G, then $p^j|[K:K \cap xHx^{-1}]$ for any x. Taking $K = S_i$, one sees that $p^j|r_i$, since the set of places where α_i has $+1$ is a union of orbits of S_i (on the columns). If $g \in G$ and $g \notin S_i$ then $g\alpha_i \neq \alpha_i$, and gS_ig^{-1} is the subgroup of G fixing $g\alpha_i$. If, moreover, gS_ig^{-1} contains a Sylow p-subgroup of S_i, then p^j divides the number of elements in each orbit (on the columns) of $S_i \cap gS_ig^{-1}$. But the set of places where α_i and $g\alpha_i$ disagree is a union of orbits of $S_i \cap gS_ig^{-1}$, so that $p^j|d$, a contradiction. Therefore, no Sylow p-subgroup of S_i is contained in a conjugate of S_i. Suppose that P is a Sylow p-subgroup of S_i (and so also of G), and that $x \in N_G(P)$, the normalizer of P. If $x \notin S_i$ then $xS_ix^{-1} \neq S_i$ but $P = xPx^{-1} \subseteq xS_ix^{-1}$, a contradiction. Hence, $N_G(P) \subseteq S_i$, and the theorem is proved.

Corollary 1. If G is a nilpotent group of collineations of M which is transitive on the columns, then either G is transitive on the rows, or one of the two orbits of G on the rows consists of one trivial row.

Proof. Unless M has only the two trivial rows, there is a prime p such that the highest power of p dividing n does not divide d. Since a Sylow p-subgroup of a nilpotent group is normal, one concludes that if G is not transitive on the rows, then G fixes a row, by Theorem 1. This proves the result.

c. Hadamard matrices and codes with two orbits. In this section there is constructed a class of Hadamard matrices with doubly transitive collineation groups; these matrices are used to obtain a class of constant-distance codes with a transitive group on the columns for which the conclusion of Corollary 1 does not hold.

Let A be the Hadamard matrix of order 4 with $+1$ on the diagonal, -1 elsewhere, and let $B = B(s)$ be the tensor product of s copies of A. The following theorem holds.

Theorem 2. For any s, the group G of collineations of $B(s)$ is doubly transitive on the columns (and also on the rows).

Proof. Denote the rows and columns of B by s-tuples, so that

$$b_{i_1, \cdots, i_s; j_1, \cdots, j_s} = a_{i_1, j_1} a_{i_2, j_2} \cdots a_{i_s, j_s}.$$

The result is obvious when $s = 1$. Suppose $s = 2$. It will be shown that the subgroup H of G fixing the column $(1, 1)$ is transitive on the remaining columns. If τ_1 and τ_2 are any permutations on four letters, then the permutation of columns sending (i_1, i_2) to $(\tau_1(i_1), \tau_2(i_2))$ is a collineation of B, sending row (i_1, i_2) to row $(\tau_1(i_1), \tau_2(i_2))$; denote this collineation by (τ_1, τ_2). It can be verified that the product of four transpositions of columns $\sigma = ((1, 4)(2, 3))((4, 1)(3, 2))((1, 3)(2, 4))((3, 1)(4, 2))$ is a collineation of B; also, $\sigma \epsilon H$. Taking σ and its products with various τ_1, τ_2, one sees that all columns other than $(1, 1)$ form a single orbit H. Moreover, some (τ_1, τ_2) moves column $(1, 1)$, so that G is transitive, and hence doubly transitive. Now suppose that $s > 2$. If τ is a collineation of $B(2)$, and if a set of two column coordinates of $B(s)$ is given, then a collineation of $B(s)$ is obtained by applying τ to the given column coordinates while keeping the remaining ones fixed. Using this type of collineation, one sees that the subgroup of G fixing column $(1, \cdots, 1)$ is transitive on the remaining columns. Hence, G is always doubly transitive on the columns and, by symmetry, also on the rows. This completes the proof.

Corollary 2. For every power 4^s of 4 ($s > 1$), there is a constant-distance code with 4^s words of length $4^s - 1$, such that the group of collineations is transitive on the columns but has two nontrivial orbits on the rows.

Proof. The matrix $B(s)$ is Hadamard, and hence its rows form a constant-distance code. Complement those rows which have a $+1$ in column $(1, \cdots, 1)$, and then delete this column. What remains is still a constant-distance code; call it C. The subgroup of G fixing $(1, \cdots, 1)$ clearly gives a group of collineations of C which is transitive on the columns. Moreover, the set of uncomplemented rows is closed under the group, so the group has two nontrivial orbits. This completes the proof.

The following remark may be made in connection with Theorem 2. Let G and H continue to have the same meanings as in Theorem 2. It follows from Corollary 1 and the proof of Corollary 2 that H is not nilpotent. However, it can actually be shown that the subgroup K

of H fixing column $(1, 2)$ is isomorphic to S_6, being generated by σ and certain (τ_1, τ_2)'s. Hence, when $s = 2$, G has order $16 \cdot 15 \cdot 720$. Also it follows that if $s > 1$, then G contains a subgroup isomorphic to S_6 which fixes $2 \cdot 4^{s-2}$ columns.

An open question: it is not known whether "nilpotent" in Theorem 1 can be replaced by "solvable."

B. Detection and Filter Theory

W. C. Lindsey, T. Nishimura, R. C. Titsworth, J. J. Stiffler, W. B. Kendall, and G. Thompson

1. Optimum Frequency Demodulation, W. C. Lindsey

a. Abstract. Design characteristics are presented for designing optimum and suboptimum frequency demodulators. Two suboptimum systems which may be implemented are considered. These are constructed by preceding the phase-locked loop, whose design is optimum for a given set of design levels, either by a bandpass limiter or an automatic gain control (AGC) amplifier. It is shown that the loop preceded by an AGC amplifier performs superior to a loop preceded by a bandpass limiter. These two system realizations are always inferior to the optimum system.

b. Introduction. The problem we wish to consider is illustrated in Fig. 1. At the transmitter a random signal $m(t)$ is used to control the instantaneous frequency of a voltage-controlled oscillator (VCO) with a gain of K_t radians/second/volt. The resulting waveform $\xi(t)$ is transmitted into the channel where it is corrupted by additive white Gaussian noise $n(t)$ with single-sided spectral density of N_0 watts/cps before being available as a waveform $\psi(t)$ to the receiver. The random processes
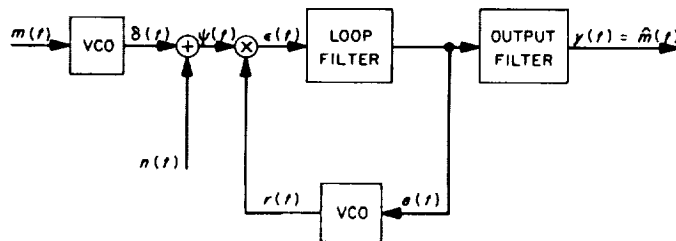

Fig. 1. Frequency modulation communication system

$m(t)$ and $n(t)$ are presumed to be wide-sense stationary and statistically independent. We intend to present the performance characteristics of the optimum phase-locked demodulator (under several restrictive assumptions) and two more easily implemented suboptimum phase-locked demodulators. The two suboptimum demodulators are obtained by fixing the pole-zero configuration of the optimum demodulators in accordance with the initial design values P, N_0, and K_t which are expected under worst channel conditions and preceding the loop by either a bandpass limiter or an automatic-gain-control amplifier.

The basic assumptions and definitions used in establishing the design techniques are as follows:

(1) The phase-locked loop is represented as a linear servo system.

(2) The threshold condition of the phase-locked loop is taken as the locus of a point which moves such that the total phase-error is always 1 rad². Further, this locus is determined using the initial design parameters P, N_0, K_t, i.e., those communication parameters under the most deleterious channel conditions expected; for example: communication at the greatest distance.

(3) The "optimum" demodulator is that filter which minimizes the mean-squared error, i.e., the difference between what appears at the output and what is desired. This is the so-called Wiener filter.

(4) The information signal $m(t)$ is considered to be a unit-variance stationary time series having the same spectral density as the output of an RC circuit possessing a 3-db frequency of a radian/second with a white noise input, i.e.,

$$S_m(s) = \frac{2s}{a^2 - s^2}.$$

(5) The output signal-to-noise ratio is defined as the reciprocal of the mean square Wiener error.

c. Receiver model. It is well known that the phase-locked loop receiver of Fig. 1, i.e., that portion of the receiver beginning with the multiplier input to the output $e(t)$, is a nonlinear device. Furthermore, the "exact" performance of the loop in the absence of modulation and additive Gaussian noise is available only in special cases.

Consequently, we seek a workable model which will closely approximate the performance of the physical device when signal modulation and noise are present. Denote the received waveform by $\psi(t)$, i.e.,

$$\psi(t) = (2P)^{1/2} \sin[\omega_0 t + \theta_1(t)] + n(t) \qquad (1)$$

where

$$\theta_1(t) = K_t \int^t m(\tau)\, d\tau$$

and (Ref. 10)

$$n(t) = 2^{1/2} n_1(t) \sin \omega_0 t + 2^{1/2} n_2(t) \cos \omega_0 t.$$

Both n_1 and n_2 are statistically independent, stationary, white, Gaussian noise processes of one-sided spectral density N_0 w/cps. The reference signal $r(t)$ at the output of the receiver's voltage-controlled oscillator is presumed to be a sinusoid whose instantaneous frequency is related to the input control voltage $e(t)$ through the relationship

$$r(t) = 2^{1/2} \sin[\omega_0 t + \theta_2(t)] \qquad (2)$$

where

$$\theta_2(t) = K_r \int^t e(\tau)\, d\tau.$$

Thus, the product of the input $\psi(t)$ and $r(t)$ may be shown to be related, in Laplace operator notation, to the phase error $\Phi(t)$ by

$$\Phi(t) = \theta_1(t) - \theta_2(t)$$

$$= \theta_1(t) - \frac{K_r}{s} F(s) \{P^{1/2} \sin \Phi(t) + n'(t)\}. \qquad (3)$$

All double frequency terms have been neglected because neither the loop filter $F(s)$ or the VCO will respond significantly to them. Further it may be shown (Ref. 10) that $n'(t)$, consisting of two noise terms, is a white Gaussian phase-noise having the same spectral density as that of the original additive process $n(t)$. Thus, to study the receiver structure from the input $\psi(t)$ to the output $e(t)$, the pertinent quantities and the phase-locked loop (PLL) may be conveniently represented by the block diagram of Fig. 2. For convenience an output filter $F_0(s)$ has also been included. Note that the receiver structure under study is placed into a form which is familiar to control engineers; we shall refer to this model as the "exact equivalent receiver." Although the loop
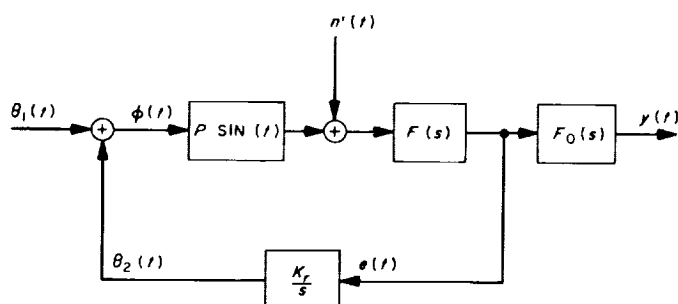
Fig. 2. Exact equivalent receiver

of Fig. 1 has been replaced by that of Fig. 2, it is still impossible to specify the "optimum" demodulator. The heart of the problem lies in the sinusoidal nonlinearity.

To linearize the loop, we make the usual assumption that $\sin \Phi \sim \Phi$; this requires that Φ be less than 1 rad. The resulting linearized loop is shown in Fig. 3. For reasons which will become obvious later, we define the following transfer functions: i.e.,

$$H(s) = \frac{y(s)}{\theta_1(s)};$$

$$H_\Phi(s) = \frac{K_r e(s)}{s\theta_1(s)} = \frac{\theta_2(s)}{\theta_1(s)}. \qquad (5)$$

And referring to Fig. 3, it is obvious that

$$F_0(s) = \frac{K_r H(s)}{s H_\Phi(s)}. \qquad (6)$$

Further, the closed-loop transfer function of the PLL may be written by inspection from Fig. 3; namely,

$$H_\Phi(s) = \frac{K_r (P)^{1/2} F(s)}{[s + K_r (P)^{1/2} F(s)]} \qquad (7)$$
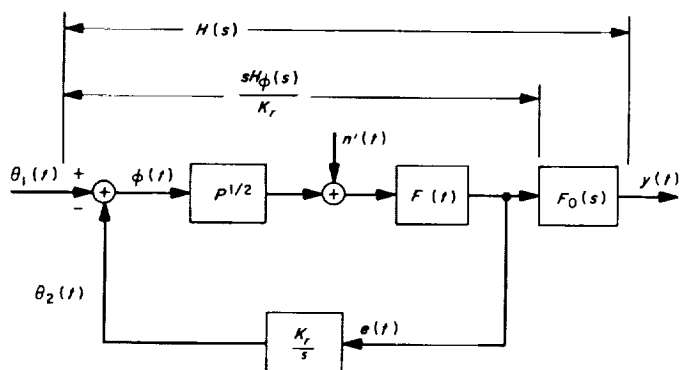

Fig. 3. Linear phase-locked receiver

and the loop filter $F(s)$, obtained from Eq. (7), is

$$F(s) = \frac{[K_{r\Phi}(P)^{1/2}]^{-1} s H_\Phi(s)}{1 - H_\Phi(s)}. \qquad (8)$$

From Eqs. (5) and (7) we see that the output $\theta_2(t)$ of the PLL is directly related to the amplitude of the input $\theta_1(t)$. Thus, the closed-loop response is a function of the *amplitude of the input signal* [Eqs. (7) and (8)]. In a physical situation where the input power may deviate from the expected value, a bandpass limiter or automatic-gain control amplifier ahead of the PLL remedies the situation.

d. Specification of the optimum filter. It may be shown (Refs. 11 and 12) that the linear filter which minimizes the mean square frequency error $\sigma_f^2 = [y(t) - m(t)]^2$ is given by

$$H(s) = \frac{a^2 (\gamma - 1)^2 s}{2K_t (s^2 + a\gamma s + a^2 \delta)} \qquad (9)$$

where

$$\gamma = (1 + 2\delta)^{1/2}; \qquad \delta = 2m_f (R)^{1/2};$$

$$m_f = K_t/a; \qquad R = P/aN_0. \qquad (10)$$

It seems appropriate to refer to m_f, the ratio of the modulation index to the 3-db frequency, as the *deviation ratio* and R as the input signal-to-noise ratio referred to a bandwidth of $a/2\pi$ cps.

Further, the filter which minimizes the mean-square phase-error ($[\sigma_\phi^2 = \theta_2(t) - \theta_1(t)]^2$ may be shown to be

$$H_\Phi(s) = \frac{a [(\gamma - 1) s + \delta a]}{s^2 + a\gamma s + a^2 \delta} \qquad (11)$$

where δ and γ are defined in Eq. (10).

The loop filter $F(s)$ may be determined using Eq. (11) in Eq. (8). Thus

$$F(s) = [P^{1/2} K_r]^{-1} \left[\frac{a (\gamma - 1) s + a (\gamma + 1)}{s + a} \right] \qquad (12)$$

whereas the output filter is determined from Eqs. (9) and (11) using Eq. (6). This filter function is

$$F_0(s) = \frac{a (\gamma - 1)^2 K_r}{2K_t [(\gamma - 1) s + a\delta]}. \qquad (13)$$

Both filters $F_0(s)$ and $F(s)$ may be realized with RC networks. It is obvious from Eqs. (12) and (13) that the optimum loop and output-filters must be adaptive, even though the frequency response equations are deceptively simple; i.e., their pole-zero configurations depend on the signal and noise-power levels. This is unfortunate in that, if the signal and noise powers are changing, the implementation of such a system would require complex servo loops and special sensors to maintain an optimum system. Consequently, in the next part of this article, we shall fix the loop and output filters to be optimum for a given set of design parameters, and see what happens to the filter performance when the conditions deviate from those expected. As a novel byproduct of the analysis, the performance of the optimum system manifests itself.

e. Specification of the suboptimum filters. As has already been pointed out, the closed-loop response of the PLL depends on the signal power P. Here we assume that the filters $F(s)$ and $F_0(s)$ are designed in accordance with some a priori selected signal and noise levels, say P and N_0, and determine the closed-loop responses $H_*(s)$ and $H(s)$ under the assumption that the received signal power is P_1 watts. Using Eqs. (12) and (13) in Eqs. (6) and (7) with P replaced by P_1 we find

$$H_*(s) = \frac{\beta[a(\gamma-1)s + a^2\delta]}{s^2 + \alpha as + a^2\delta\beta} \tag{14}$$

$$H(s) = \frac{\beta a^2(\gamma-1)a}{2K_t[s^2 + \alpha as + a^2\delta\beta]} \tag{15}$$

where

$$\alpha = 1 + \beta(\gamma - 1)$$

$$\beta = \left(\frac{P_1}{P}\right)^{1/2}. \tag{16}$$

If $P_1 = P$, Eqs. (14) and (15) reduce to Eqs. (9) and (10) as they should.

f. The output signal-to-noise ratio. The output signal-to-noise ratio is defined here as the ratio of the mean-squared signal power to the mean-squared value of the frequency error, i.e.,

$$\rho_0 = \frac{\overline{[m(t)]^2}}{\sigma_f^2} = \frac{1}{\sigma_f^2}. \tag{17}$$

The mean-square frequency-error is found from the following relationship,

$$\sigma_f^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} S_{\theta_1}(s)|L(s) - H(s)|^2 ds$$

$$+ \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} S_{n'}(s)|H(s)|^2 ds \tag{18}$$

where $S_{\theta_1}(s)$ is the spectral density of the signal, $L(s)$ is the desired linear operation of the input and $S_{n'}(s)$ is the spectral density of the additive noise. Eq. (18) is a standard equation, the derivation of which is immediate when the signal and noise are transformed into their Laplace notation. The mean-square error is then found by squaring the absolute difference between the actual output signal and the desired output signal and integrating over all frequencies. From Eq. (13), we see that the mean-squared error is composed of two parts. The first part is due to the fact that the transfer function differs from $L(s)$. We shall refer to this portion of the error as the signal distortion, i.e.,

$$S_d = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} S_{\theta_1}(s)|L(s) - H(s)|^2 ds. \tag{19}$$

The other portion of the error is due to the additive phase-noise. First, we compute the signal distortion term using Eq. (15) and the signal-phase spectral density in Eq. (19) with $L(s) = s/K_t$. Carrying out the necessary integration, we find

$$S_d = \frac{4\delta\beta(\delta+\alpha) + 4\delta\beta(\alpha^2 - K) + K^2(\alpha+1)}{4\delta\beta[(\alpha+1)(\delta+\alpha) - \delta\beta]} \tag{20}$$

where

$$\alpha = 1 + \beta(\gamma - 1)$$

$$K = \beta(\gamma^2 - 1) - (\gamma - 1)^2.$$

On the other hand, the variance of the phase-noise error is easily found by substituting into the second term of Eq. (18) and integrating. The result is

$$\sigma_{n'}^2 = \frac{N_1}{N_0}\frac{(\gamma-1)^2}{\alpha(\gamma+1)^2} \tag{21}$$

where N_1 is the new noise spectral density. Hence, the new output signal-to-noise ratio ρ_0 is, from Eqs. (20) and (21),

$$\rho_0 = \frac{1}{S_d + \frac{N_1}{N_0}\left[\frac{(\gamma - 1)^2}{\alpha(\gamma + 1)^2}\right]} \tag{22}$$

which is the required result.

The new phase error σ_1^2 resulting from a change in received signal power may be written

$$\sigma_1^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} S_{\theta_1}(s)|1 - H_\Phi(s)|^2 ds$$

$$+ \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} S_{n'}(s)|H_\Phi(s)|^2 ds. \tag{23}$$

This becomes

$$\sigma_1^2 = \frac{2m_f^2}{\alpha\beta(\gamma^2 - 1)}\left[1 + \frac{N_1}{N_0}\left(\frac{2\beta(\gamma - 1)^2 + \gamma^2 - 1}{(\gamma^2 - 1)}\right)\right] \tag{24}$$

when Eq. (14) and the appropriate signal and noise spectral densities are substituted into Eq. (23), and the integration performed. Thus, Eqs. (22) and (24) are the results needed for specifying the receiver performance. If $N_1 = N_0$, $P_1 = P$ Eq. (24) becomes

$$\sigma_\Phi^2 = \sigma_1^2 = \frac{8m_f^2}{(\gamma + 1)^2(\gamma - 1)}. \tag{25}$$

This is the equation specifying the threshold locus. If $\beta = 1$, $N_1 = N_0$, Eq. (22) gives the output signal-to-noise ratio for the optimum system.

g. *Output signal-to-noise ratio when the loop is preceded by an automatic gain control amplifier.* In this case, we assume that the loop is preceded by an automatic-gain-control amplifier to hold the signal power constant, i.e., $P = P_1$ for all variations. In this case $\beta = 1$ and S_d reduces to

$$S_d = \frac{3\gamma - 1}{\gamma(\gamma + 1)}.$$

Thus, Eq. (22) may be rewritten as

$$\rho_0 = \frac{1}{\frac{3\gamma - 1}{\gamma(\gamma + 1)} + \frac{R_D}{R}\left[\frac{(\gamma - 1)^2}{\gamma(\gamma + 1)^2}\right]} \tag{26}$$

where

$$R_D = \frac{P}{aN_0}; \qquad R = \frac{P}{aN_1}.$$

The value R_D may be considered to be the input design signal-to-noise ratio referred to a bandwidth of $a/2\pi$ cps and R the actual input signal-to-noise ratio referred to the same bandwidth.

h. *Output signal-to-noise ratio when the loop is preceded by a bandpass limiter.* In this case we need to relate N_1/N_0 and P_1/P of Eqs. (20) and (21) to the input-output properties of a limiter followed by a bandpass filter. (A bandpass limiter is an electronic device which consists of a limiter followed by a bandpass filter.)

Davenport (Ref. 13) has shown that

$$\left(\frac{S}{N}\right)_o \approx 2\left(\frac{S}{N}\right)_i \qquad \text{for}\left(\frac{S}{N}\right)_i \to \infty$$

$$\left(\frac{S}{N}\right)_o \approx \frac{\pi}{4}\left(\frac{S}{N}\right)_i \qquad \text{for}\left(\frac{S}{N}\right)_i \to 0 \tag{27}$$

where

$\left(\dfrac{S}{N}\right)_o$ = signal-to-noise power ratio at the limiter output

$\left(\dfrac{S}{N}\right)_i$ = signal-to-noise power ratio at the limiter input

$$S_o + N_o \approx \frac{2\alpha^2}{\pi}\left[\frac{4}{\pi} + \left(\frac{S}{N}\right)_i\right] \qquad \text{for}\left(\frac{S}{N}\right)_i \to 0$$

$$S_o + N_o \approx \frac{2\alpha^2}{\pi}\left[\frac{4}{\pi} + \left(\frac{N}{S}\right)_i\right] \qquad \text{for}\left(\frac{S}{N}\right)_i \to \infty \tag{28}$$

Further,

$$\frac{\pi}{4}\left(\frac{S}{N}\right)_o\left(\frac{S}{N}\right)_i^{-1} \leq 2, \tag{29}$$

a variation of about 4.06 db. Eqs. (28) and (29) suggest that the bandpass limiter is capable of holding the output power relatively constant.[1] Consequently, if the noise input to the system increases, the signal strength at the limiter output is forced to decrease. However, the loop noise bandwidth B_L, i.e.,

$$2B_L = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} |H_*(s)|^2 \, ds$$

is directly dependent on the signal power [Eq. (7)]. Thus, an increase in input noise, for a fixed signal power, causes the output signal power to be reduced, hence the loop-bandwidth; and hence the phase jitter at the VCO output will be a smaller percentage of the input noise. Therefore,

the over-all demodulator appears to be self-adaptive. Using Eqs. (28) and (29) and assuming that the gain of the bandpass limiter is adjusted such that the output signal-power-plus-noise-power are those values for which the optimum loop was designed, it is easy to show that

$$\frac{N_1}{N_0} = \frac{1 + R_D}{1 + R}$$

$$\frac{P_1}{P} = \frac{R}{R_D}\left(\frac{1 + R_D}{1 + R}\right) \tag{30}$$

where

R = input signal-to-noise ratio of the bandpass limiter

R_D = input signal-to-noise ratio for which the loop was designed

---

[1]That this is indeed true has been experimentally established by Youla of this Laboratory. An exact plot of the output signal and noise powers is given in Ref. 13.
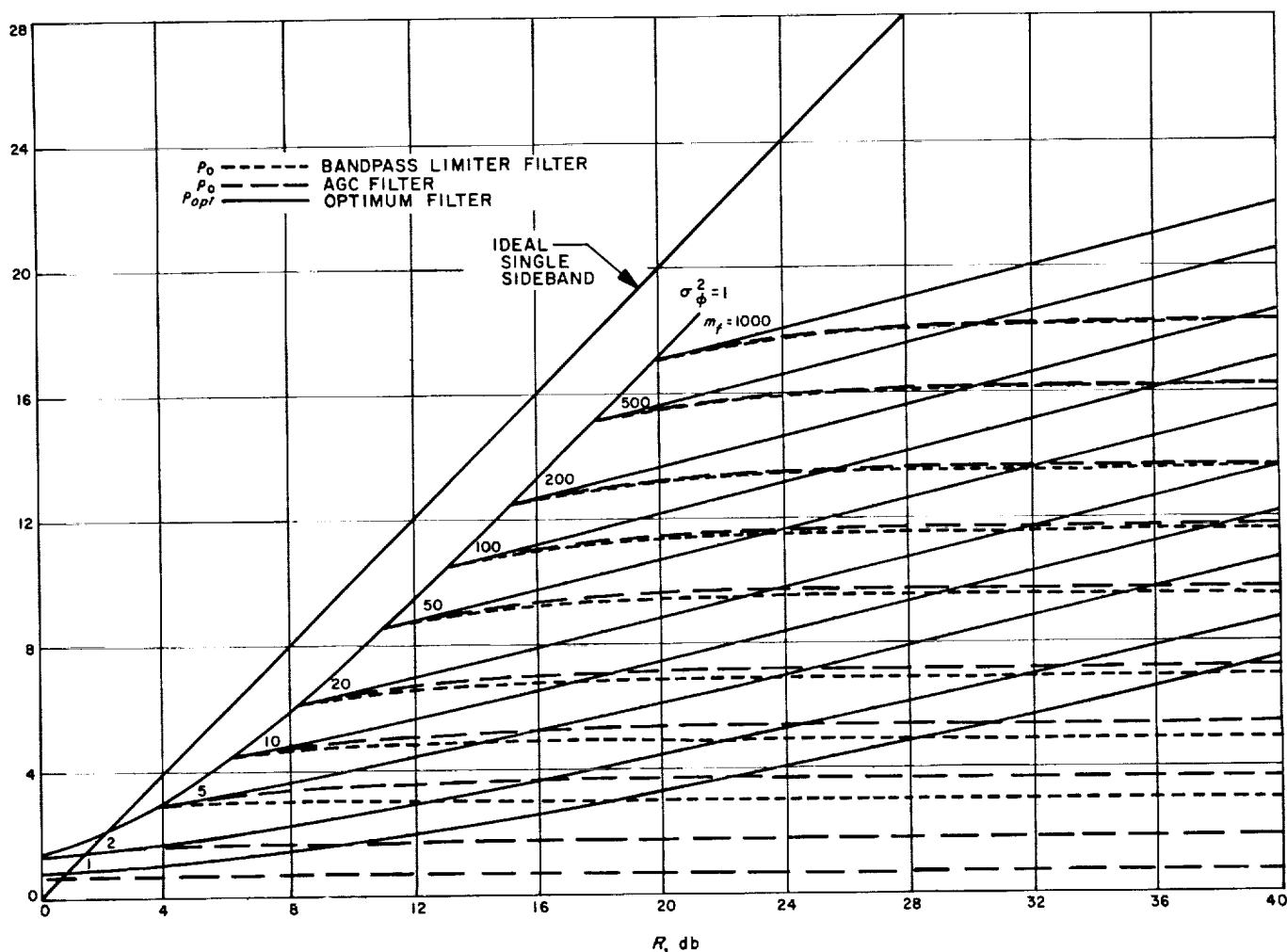

Fig. 4. Optimum and suboptimum system performance characteristics

P_1, N_1 = bandpass limiter output noise and signal levels, respectively

P, N_0 = signal and noise levels for which the loop was designed

$$\rho_0 = \frac{1}{S_d + \left(\dfrac{1 + R_D}{1 + R}\right)\left[\dfrac{(\gamma - 1)^2}{\alpha(\gamma + 1)^2}\right]} \quad (31)$$

where S_d is given by Eq. (20) and

$$\alpha = 1 + \beta(\gamma - 1)$$

$$k = \beta(\gamma^2 - 1) - (\gamma - 1)^2$$

$$\beta = \left[\frac{R}{R_D}\left(\frac{1 + R_D}{1 + R}\right)\right]^{\frac{1}{2}}. \quad (32)$$

The corresponding system phase-error is given by Eq. (24) where α, β, and N_1/N_0 are defined in Eq. (32).

i. System design curves and comparison. Shown in Fig. 4 are graphical results relating the output signal-to-noise ratio ρ to the input signal-to-noise ratio R referred to a bandwidth of $a/2\pi$ cps for various values of the deviation ratio m_f. It is clear that the optimum system always outperforms either of the more easily implemented suboptimum systems. Also, note that the AGC system is superior to the bandpass limiter system, but the amount of this superiority becomes imperceptible as m_f increases. It is also evident from Fig. 4 that, once the system has been optimized with respect to a given set of design parameters, not much is gained by increasing the signal-to-noise ratio R from the design level. The reason for this is that the signal distortion does not diminish.

For comparison purposes, we have plotted in Fig. 5 the ratio ρ_o ρ_{opt}, i.e., the ratio of the output signal-to-noise
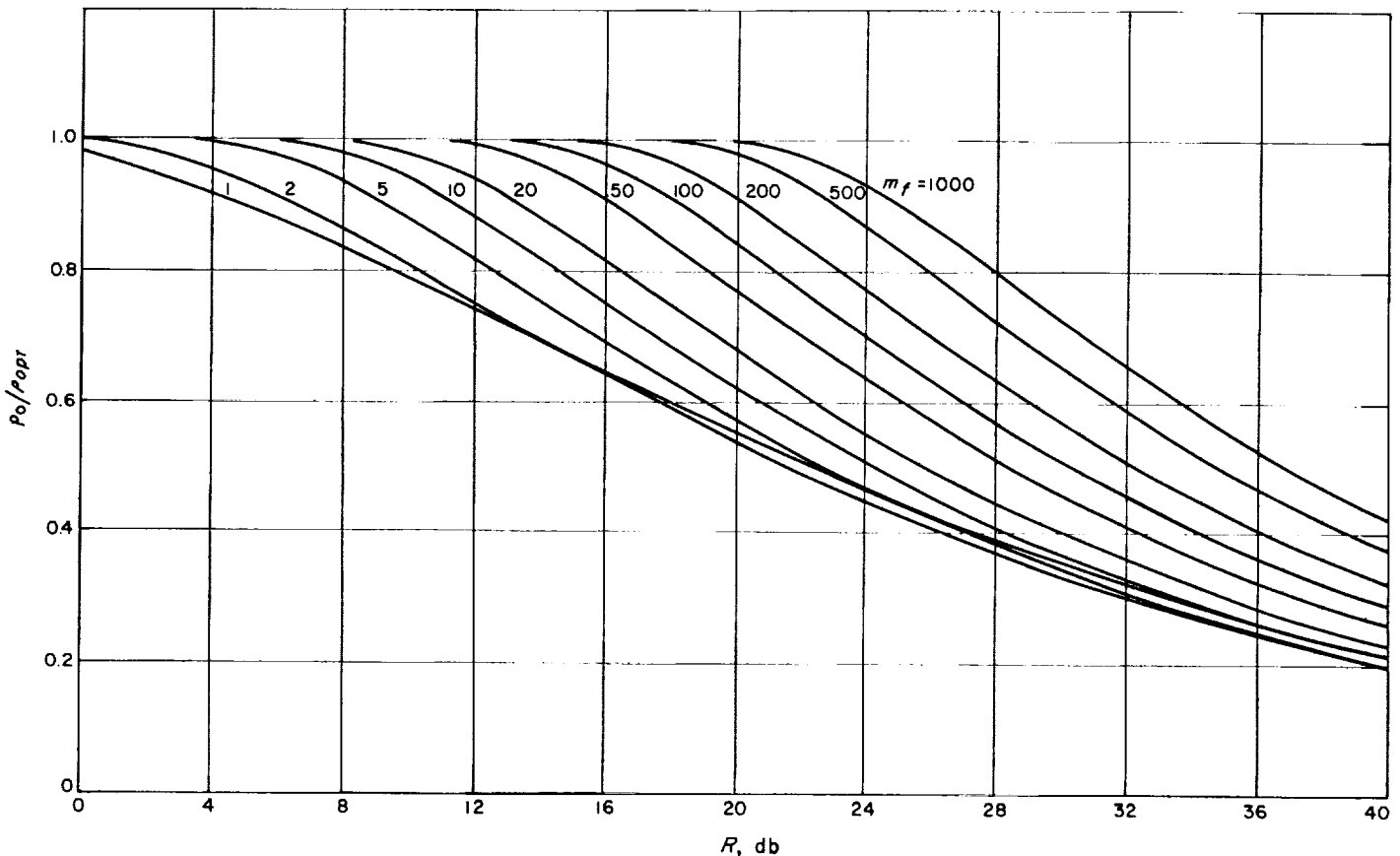

Fig. 5. System comparison

ratio for the optimum system to the output signal-to-noise ratio of the AGC system. For values of R relatively close to threshold the AGC system performs almost optimally.

Finally, we have indicated in Fig. 4 the "input–output signal-to-noise ratio" characteristic for an ideal-single-sideband-suppressed carrier-system" transmitting a steady tone. Notice that the threshold characteristic of the phase-locked receiver intersects the "ideal-single-sideband" system at an input signal-to-noise ratio of approximately $R = 2$ db. For $R > 2$ db, the single-sideband system performance increases (for the range considered) to a maximum of about 3 db better than the phase-locked receiver. In the past, this line has been used as means of comparison; however, a more legitimate and meaningful comparison could be made by employing a phase-locked receiver which is "amplitude matched" to the spectrum considered here. A reasonable conjecture would be (all things being equal) that the single-sideband phase-locked receiver would be inferior to the FM system when R is large. This relationship is currently being studied.

In the past, several authors (Ref. 14) have compared FM, AM, and PM systems under the assumption that the signaling waveform is monochromatic and the additive disturbance is Gaussian. Their comparison is also made on the basis of input-versus-output signal-to-noise ratios, but their definition of the output signal-to-noise ratio differs from that considered here. The major difference is that the signal distortion term in Eq. (19) is removed from the denominator and placed in the numerator. Consequently, as the "input signal-to-noise ratio" increases indefinitely so does the "output signal-to-noise ratio." Here we found for the suboptimum systems that, as R increases indefinitely, the output signal-to-noise ratio is determined solely by the signal-distortion designed into the system. For purposes of this study, this viewpoint appears to be the more definitive of the two. Further, we are speaking about an optimum form of demodulation, whereas the Foster-Seeley detector is truly suboptimum, and any time a comparison is to be made between systems, extreme care must be exercised in making sure that the definitions are consistent between systems; hence the conclusions reached are legitimate. When one is comparing optimum systems no such trouble arises. This is the primary reason that the "amplitude matched" single-sideband phase-locked modulation system would be more appropriate system to compare with the FM system under consideration.

2. Design of Phase-Locked Loop Systems With Correlated Noise Input, T. Nishimura

a. Introduction. The performance of the phase-locked loop system is analyzed when three types of controllers, namely, proportional (P), proportional and differential (PD) and proportional, integral and differential (PID) controllers are applied to the feed-back loop.

The performance index J is defined as the mean-squared value of the deviation between the signal estimate and the signal itself, and is expressed as the sum of the mean-squared control error over the observation time T_o and the noise variance.

Performance criteria in this analysis are, then, to minimize J, or equivalently, to maximize the output signal-to-noise ratio X_0 by choice of the control. When the input noise (phase jitter) is white, this optimum for a step signal input is a proportional (gain) controller (Ref. 15). However, when the phase noise is correlated (for example, when it is approximated by an RC filtered white noise) the above is not true.

In the case of digital phase-locked loop systems employing PN codes as the modulating signals of RF carriers, the input noise to the loop is expected to be band-limited by the low-pass filter in the demodulator preceding the limiter. Thus, it becomes necessary to introduce design criteria under the assumption of correlated noise rather than the conventional assumption of white noise. A qualitative comparison of control effects of P controller, PD controller, and RC lead network is performed in terms of output signal-to-noise ratio X_0, and it will be shown that the PD controller is superior to the P controller when the input noise is correlated.

The simplified phase-locked loop system described as a linear feedback system is shown in Fig. 6. The correlator of the input PN codes and local ones is assumed to be operated within the linear region. The signal $s(t)$ is then the phase of the incoming PN. The VCO is represented by an integrator. The correlator gain, VCO gain, and the
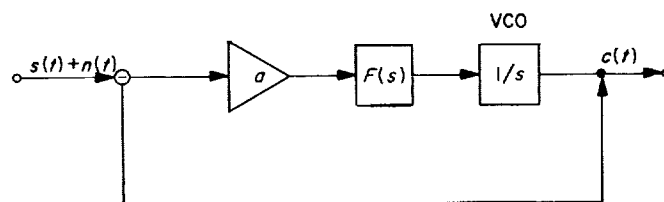

Fig. 6. Block diagram of phase-locked loop system

amplifier gain are combined and represented by an adjustable gain a. Also, a compensating filter $F(s)$ is inserted.

The open-loop transfer function is

$$GF(s) = \frac{a}{s} F(s) \qquad (1)$$

and the over-all transfer function is

$$K(s) = \frac{GF(s)}{1 + GF(s)} = \frac{aF(s)}{s + aF(s)}. \qquad (2)$$

b. Proportional (gain) control. When $F(s)$ is equal to 1 in Fig. 6 and a signal $s(t)$ and a noise $n(t)$ are supplied to the system, the output $c(t)$ of the system consists of the signal component $c_s(t)$ and the noise component $c_n(t)$, corresponding to $s(t)$ and $n(t)$ at the input. The deviation $e(t)$ is defined as the difference between the input signal $s(t)$ and the output $c(t)$

$$\begin{aligned} e(t) &= s(t) - c(t) \\ &= s(t) - c_s(t) - c_n(t) \\ &= e_s(t) - c_n(t) \end{aligned} \qquad (3)$$

where

$$e_s(t) = s(t) - c_s(t). \qquad (4)$$

This is a deterministic process, namely the transient phase error.

The performance index J is chosen as the time average of the ensemble average of $e^2(t)$. If the input noise has zero mean, and has no correlation with input signals, J is reduced as follows:

$$J = \langle \overline{E^2} \rangle_{T_o} = \langle \overline{e_s^2} \rangle_{T_o} + \langle \overline{c_n^2} \rangle_{T_o} \qquad (5)$$

where the bar indicates the ensemble average and $\langle \ \rangle$ indicates the time average over a finite observation time T_o.

If T_o is sufficiently large, the upper limit T_o of integration of $e_s^2(t)$ can be replaced by ∞ as $e_s^2(t)$ becomes negligibly small for large t, because we have assumed only transient error in the loop. Also, for large T_o, the ensemble average and time average of c_n^2 become equal

and approach the infinite time average $\langle c_n^2 \rangle_\infty$. Thus, Eq. (5) for large T_o is approximately

$$J = \frac{1}{T_o} \int_0^\infty e_s^2(t)\, dt + \langle c_n^2 \rangle_\infty. \qquad (6)$$

When Parseval's theorem is applied to the above equation, J can be described by an integral in the frequency domain. Let $\Phi_{c_n c_n}(s)$ and $\Phi_{nn}(s)$ be power spectral densities of $c_n(t)$ and $n(t)$ and $E_s(s)$ and $S(s)$ be the Laplace transforms of $e_s(t)$ and $s(t)$, respectively. Then

$$\begin{aligned} J &= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left[\frac{1}{T_o} E_s(s) E_s(-s) + \Phi_{c_n c_n}(s) \right] ds \\ &= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left[\frac{1}{T_o}(1 - K(s))(1 - K(-s)) S(s) S(-s) \right. \\ &\qquad \left. + K(s) K(-s) \Phi_{nn}(s) \right] ds. \end{aligned} \qquad (7)$$

The design objective of this analysis is to minimize J by constructing an optimum transfer function $K_0(s)$. In the case of a P controller, $K_0(s)$ is derived by finding an optimum gain a_0.

When the input signal is a step function of magnitude Q

$$S(s) = \frac{Q}{s} \qquad (8)$$

and the input noise is approximated by RC-filtered white noise of (double-sided) spectral density Φ_N:

$$\Phi_{nn}(s) = \frac{\Phi_N}{(1 + \tau s)(1 - \tau s)}. \qquad (9)$$

J of Eq. (7) is computed by evaluating the contour integral on the left half s-plane

$$\begin{aligned} J &= \frac{Q^2}{2aT_o} + \frac{a\Phi_N}{2(1 + \tau a)} \\ &= Q^2 \left[\frac{1}{2aT_o} + \frac{\tau a}{X_I(1 + \tau a)} \right]. \end{aligned} \qquad (10)$$

X_I is the signal-to-noise ratio (SNR) of the input defined as follows:

$$X_I = \frac{Q^2}{P_N} = \frac{2\tau Q^2}{\Phi_N} \qquad (11)$$

and P_N is the input noise power

$$P_N = \frac{\Phi_N}{2\tau}. \qquad (12)$$

Differentiating J of Eq. (10) by a and equating it to zero yields an optimum a_0 which minimizes J.

$$a_0 = \frac{\left(\frac{X_I}{2\tau T_o}\right)^{\frac{1}{2}}}{1 - \left(\frac{\tau X_I}{2T_o}\right)^{\frac{1}{2}}} \qquad \text{for } X_I < \frac{2T_o}{\tau} \qquad (13)$$

When $X_I > 2T_o/\tau$, J becomes a monotonic decreasing function of a and approaches its minimum value as a goes to infinity. This is the case when the input signal-to-noise ratio is very large or the input noise is strongly correlated. And it becomes more effective to reduce signal error by increasing gain rather than attempting noise reduction. Thus, the infinite gain becomes the optimum condition that reduces signal error to zero, leaving the noise power P_N unfiltered.

Such an infinite increase of gain is impractical; however, this result is presented as a limiting case only. In practice, the maximum gain that can be attained by the amplifier is best, under the condition $X_I > 2T_o/\tau$.

The mean-squared deviation J with these optimum gains are computed

$$J = \begin{cases} Q^2\left[\left(\frac{2\tau}{T_o X_I}\right)^{\frac{1}{2}} - \frac{\tau}{2T_o}\right] & \text{for } X_I < \frac{2T_o}{\tau} \\ \\ \frac{Q^2}{X_I} = P_N & \text{for } X_I > \frac{2T_o}{\tau} \end{cases} \qquad (14)$$

The output SNR X_{0g} is defined as the ratio of Q^2 and J and it is

$$X_{0g} = \frac{Q^2}{J} = \begin{cases} \dfrac{1}{\left(\dfrac{2\tau}{T_o X_I}\right)^{\frac{1}{2}} - \dfrac{\tau}{2T_o}} & \text{for } X_I < \dfrac{2T_o}{\tau} \\ \\ X_I & \text{for } X_I > \dfrac{2T_o}{\tau} \end{cases} \qquad (15)$$

The above relation is normalized by letting

$$\beta = \left(\frac{2T_o}{\tau}\right)^{\frac{1}{2}} \qquad (16)$$

$$x = \frac{X_I}{\beta^2} \qquad (17)$$

$$y_{0g} = \frac{X_{0g}}{\beta^2} \qquad (18)$$

The normalized output SNR y_{0g} becomes

$$y_{0g} = \frac{x^{\frac{1}{2}}}{2 - x^{\frac{1}{2}}} \qquad \text{for } x < 1$$
$$\quad = x \qquad\qquad \text{for } x > 1 \qquad (19)$$

and y_{0g} is plotted in Fig. 7 as a function of x.

c. Proportional and differential control. The proportional (gain) control described in the previous part of this article has a certain limitation. For example, as observed from Eq. (13), the optimum gain a_0 becomes very large as X_I approaches $2T_o/\tau$. Such an increase of gain may be restricted by the saturating characteristics of amplifier or by the nonlinear characteristics of VCO.

In such a case, the compensation by means of network is more suitable than the gain adjustment. In this part of this article, such an optimum filter that minimizes J is derived, and its function is compared with that of the gain controller. Wiener's conventional optimization method is applied for this purpose and an optimum transfer function $K_0(s)$ can be derived as follows:

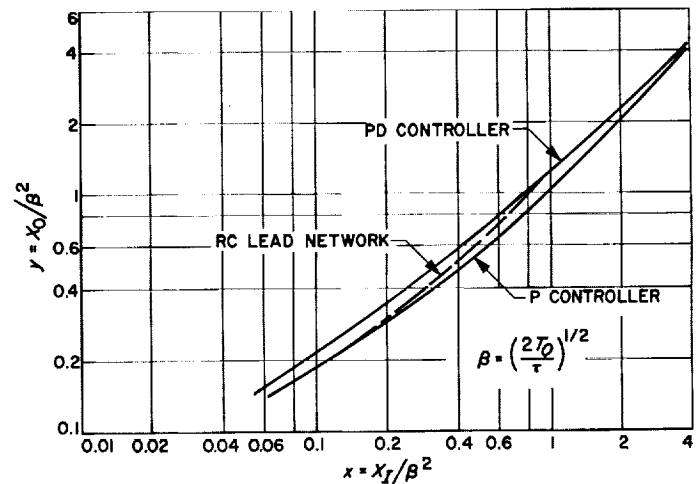$$K_0(s) = \frac{1}{Z(s)S(s)}\left[\frac{S(s)}{Z(-s)}\right]^+, \qquad (20)$$

Fig. 7. Performances of P, PD and RC lead network controller under correlated noise input

$Z(s)$ is obtained by the following factorization:

$$Z(s)Z(-s) = 1 + \frac{T_o \Phi_{nn}(s)}{S(s)S(-s)}, \qquad (21)$$

in which $Z(s)$ is factored in such a way that all its zeroes and poles are on the left half s-plane. The operation indicated by $[\;\;]^+$ is to expand its content by means of partial fractions and then to select all the terms having left half plane poles.

The optimum filter is obtained by

$$F_0(s) = \frac{1}{G(s)} \frac{K_0(s)}{1 - K_0(s)}. \qquad (22)$$

When the input signal is a step function of magnitude Q and $\Phi_{nn}(s)$ is given by Eq. (9), $Z(s)$ is derived from Eq. (21)

$$Z(s) = \frac{1 + \alpha s}{1 + \tau s} \qquad (23)$$

where

$$\alpha = \left(\tau^2 + \frac{2\tau T_o P_N}{Q^2} \right)^{\frac{1}{2}} = \tau \left(1 + \frac{\beta^2}{X_I} \right)^{\frac{1}{2}} \qquad (24)$$

$K_0(s)$ is obtained from Eq. (20)

$$K_0(s) = \frac{1 + \tau s}{1 + \alpha s}. \qquad (25)$$

Then $F_0(s)$ is derived from Eq. (22)

$$F_0(s) = \frac{1 + \tau s}{a(\alpha - \tau)}. \qquad (26)$$

This $F_0(s)$ consists of a constant term (proportional gain) and a differentiating operator. This is a typical PD (proportional and differential) controller. This "lead" network makes the system have a fast response for changes in the input signal, and it also reduces the high-frequency component of noise by a factor τ/α. With this optimum filter $F_0(s)$, the deviation J reduces to

$$J = Q^2 \left[\left(\frac{\tau^2}{T_o^2} + \frac{2\tau}{X_I T_o} \right)^{\frac{1}{2}} - \frac{\tau}{T_o} \right]. \qquad (27)$$

The output SNR ratio is

$$X_{0s} = \frac{Q^2}{J} = \frac{1}{2} X_I + \frac{1}{2} \left(X_I^2 + \frac{2T_o X_I}{\tau} \right)^{\frac{1}{2}}. \qquad (28)$$

This is normalized by the relations of Eq. (16) through (18)

$$y_{0s} = \frac{X_{0s}}{\beta^2} = \frac{x}{2} + \frac{1}{2} [(1 + x)x]^{\frac{1}{2}}. \qquad (29)$$

d. Compensation by RC lead network. A simple RC lead network as shown in Fig. 8 may be used for the dual purpose of reducing the output noise power and making the system respond quickly to input variation. The transfer function of this lead network is

$$F(s) = \frac{m(1 + \tau_I s)}{1 + m\tau_I s} \qquad (30)$$

where

$$\tau_I = CR_1 \qquad (31)$$

$$m = \frac{R_2}{R_1 + R_2} < 1. \qquad (32)$$

τ_I is equated to τ in order to remove the pole in the input noise spectrum. Then the over-all transfer function $K(s)$ is described as follows

$$K(s) = \frac{am(1 + s\tau)}{ma + (1 + ma\tau)s + m\tau s^2} \qquad (33)$$

from which J of Eq. (7) is computed.

$$J = \frac{1}{2(1 + am\tau)} \left[\frac{1 + am^2\tau}{T_o ma} + \frac{2\tau ma}{X_I} \right]. \qquad (34)$$
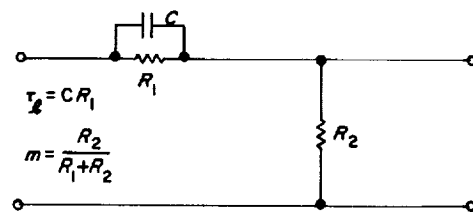

Fig. 8. RC lead network

Differentiating this with respect to m and equating the result to zero yields an opitmum ratio of resistances

$$m_0 = \frac{1}{a_T \left(1 + \frac{2aT_o}{X_I}\right)} \left\{ a_T + \left[a^2_T + a_T \left(1 + \frac{2aT_o}{X_I}\right) \right]^{\frac{1}{2}} \right\}. \tag{35}$$

And the optimum SNR ratio of the output is obtained by substituting the above m_0 into Eq. (34)

$$X_{0l} = \frac{Q^2}{J} = am_0T_o. \tag{36}$$

Since X_{0l} is an increasing function of a, larger gain a yields better X_{0l}, provided the open-loop gain a is adjustable. As a limit, X_{0l} approaches X_{0s} given by Eq. (28) when a goes to infinity. And the form of $aF(s)$ approaches the form of PD controller $aF_0(s)$ of part c since $m \to 0$, $am_0 \to (1/\alpha - \tau)$ as $a \to \infty$, and thus $aF(s) \to (1 + \tau s)(\alpha - \tau) = aF_0(s)$. Although the performance of this lead network is slightly inferior to that of $F_0(s)$ (PD controller) its advantage lies in the fact that it is realizable by a simple RC network. Also its performance is superior to that of the proportional (gain) controller.

When the open-loop gain a is equated to the optimum gain of P controller given by Eq. (13) for the sake of comparison, the normalized output SNR y_{0l} under the lead network compensation can be reduced to

$$y_{0l} = \frac{X_{0l}}{\beta^2} = \frac{x^{\frac{1}{2}}}{2\left[(1 - x^{\frac{1}{2}})^{\frac{1}{2}} - x^{\frac{1}{2}}\right]} \quad \text{for } x < 1. \tag{37}$$

For $x > 1$, a_0 becomes infinite and the lead network controller becomes identical with the PD controller as discussed in the above. Hence

$$y_{0l} = y_{0s} \quad \text{for } x > 1. \tag{38}$$

The normalized output SNR's for three types of controllers, namely proportional, proportional and differential and RC lead network are plotted in Fig. 7 as functions of the normalized input SNR. The figure shows that a PD controller demonstrates better performance than a P controller by about 1 db over the range indicated, and RC lead network lies in between.

e. Proportional, integral, and differential (PID) control. When the input is a phase ramp-function, it is well-known that the compensated open-loop transfer function $GF(s)$ must have double poles at the origin in order to

reduce the steady state error to zero. This implies that $F(s)$ has a single pole at the origin. Also it is expected that $GF(s)$ have a zero at $-1/\tau$ in order to remove the pole at $-1/\tau$ of the correlated noise input.

The desired optimum filter is also described by Eq. (20) of part c. Let the signal input be

$$S(s) = \frac{P}{s^2} \tag{39}$$

and the noise be given by Eq. (9). Then, from Eq. (21)

$$Z(s) = \frac{1 + \delta s + \gamma s^2}{1 + \tau s} \tag{40}$$

where

$$\gamma = \left(\frac{T_o \Phi_N}{P^2}\right)^{\frac{1}{2}} \tag{41}$$

and

$$\delta = (\tau^2 + 2\gamma)^{\frac{1}{2}}. \tag{42}$$

Then $K_0(s)$ is derived from Eq. (20)

$$K_0(s) = \frac{(1 + \tau s)(1 + (\delta - \tau)s)}{1 + \delta s + \gamma s^2}. \tag{43}$$

Thus, the optimum filter is obtained

$$F_1(s) = \frac{(1 + \tau s)(1 + s(\delta - \tau))}{a(\gamma - \delta\tau + \tau^2)s}$$

$$= \frac{1 + \delta s + (\delta - \tau)\tau s^2}{a(\gamma - \delta\tau + \tau^2)s}. \tag{44}$$

As observed from the above equation, $F_1(s)$ is a typical PID controller and possesses the properties that were stated at the beginning of this article, having single integrator and a zero at $-1/\tau$ to remove the pole of noise spectral density.

3. A Theorem on Low-Pass Filtering, R. C. Titsworth

Let $v(t)$ be a stationary random process with mean μ. By inserting $v(t)$ into a low-pass filter, we wish to estimate μ. Let us approximate the low-pass filter by an integrator:

$$w(t) = \frac{1}{T} \int_{t-T}^{t} v(\zeta) \, d\zeta.$$

Then, clearly, the mean output is μ:

$$\mu = E\left[w\left(t\right)\right].$$

Also, it follows that the variance σ_w^2 of the output is

$$\sigma_w^2 = \frac{1}{T^2} \int_{t-T}^{t} \int_{t-T}^{t} E\left[v\left(\zeta\right)v\left(\xi\right)\right] d\zeta \, d\xi - \mu^2$$

$$= \frac{1}{T^2} \int_{t-T}^{t} \int_{t-T-\zeta}^{t-\zeta} E\left[v\left(\zeta\right)v\left(\zeta+\tau\right)\right] d\tau \, d\zeta - \mu^2.$$

Since the process is stationary, σ_w^2 is independent of t, so

$$\sigma_w^2 = \frac{1}{T^2} \int_{0}^{T} \int_{-t}^{T-t} R_v\left(\tau\right) d\tau \, dt - \mu^2$$

$$= \frac{2}{T^2} \int_{0}^{T} \left(T - \tau\right) R_v\left(\tau\right) d\tau - \mu^2$$

where $R_v\left(\tau\right)$ is the autocorrelation function of $v\left(t\right)$. The second form of this equation is obtained by interchanging orders of integration (Fig. 9).

Let us now note that $R_v\left(\tau\right)$ can be separated into three parts: a part $R_p\left(\tau\right)$ giving rise to spectral spikes at frequencies $|f| > 0$, an aperiodic part $R_u\left(\tau\right)$, and the value μ^2.

$$R_v\left(\tau\right) = R_p\left(\tau\right) + R_u\left(\tau\right) + \mu^2.$$

$R_p\left(\tau\right)$ is composed of periodic components, and for T sufficiently large, we shall show its contribution is arbitrarily small. $R_u\left(\tau\right)$ represents a process having a spectral density, so

$$\lim_{\tau \to \infty} R_u\left(\tau\right) = 0.$$
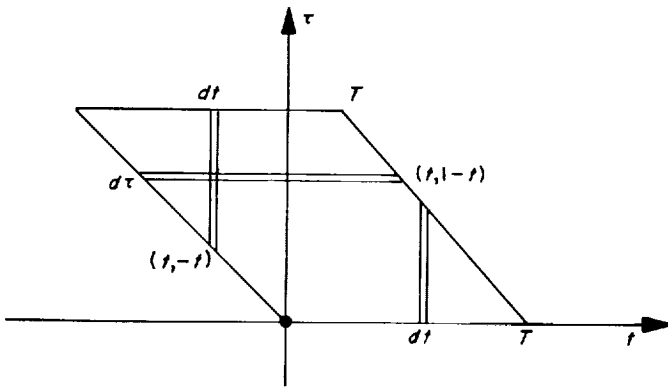
Fig. 9. Interchanging the order of integration

For our applications, we assume $R_u\left(\tau\right)$ is absolutely integrable.

If

$$R_p\left(\tau\right) = \sum_{n=-\infty}^{+\infty} a_n e^{j\omega_n \tau},$$

its contribution σ_{wp}^2 to σ_w^2 is

$$\sigma_{wp}^2 = \frac{1}{T^2} \int_{0}^{T} \int_{-t}^{T-t} R_p\left(\tau\right) d\tau \, dt$$

$$= \frac{1}{T^2} \sum_{n=-\infty}^{+\infty} \frac{a_n}{\omega_n^2} \left(e^{j\omega_n T} - 1\right)\left(e^{-j\omega_n T} - 1\right)$$

$$= \frac{4}{T^2} \sum_{n=-\infty}^{+\infty} \frac{a_n}{\omega_n^2} \sin^2 2\omega_n T$$

and therefore

$$\sigma_{wp}^2 \leq \frac{4}{T^2} \sum_{n=-\infty}^{+\infty} \frac{a_n}{\omega_n^2} = \frac{K}{T^2}$$

for some constant K.

Let $\alpha = \alpha\left(\epsilon\right)$ be that value of τ beyond which

$$\int_{0}^{\alpha} \left| R_u\left(\tau\right)\right| d\tau = B < \infty;$$

$$\int_{\alpha}^{\infty} \left| R_u\left(\tau\right)\right| d\tau = \epsilon$$

for any arbitrarily chosen small value ϵ. Define

$$\beta = \int_{0}^{\alpha} R_u\left(\tau\right) d\tau.$$

The form of $R_u\left(\tau\right)$ will appear as in Fig. 10. The shape of $R_u\left(\tau\right)$ on $\left(0, \alpha\right)$ is immaterial, and we note

$$R_u\left(\tau\right) \leq R_u\left(0\right) = \sigma_v^2 - R_p\left(0\right).$$

The contribution of $R_u\left(\tau\right)$ to σ_w^2 is σ_{wu}^2, given by

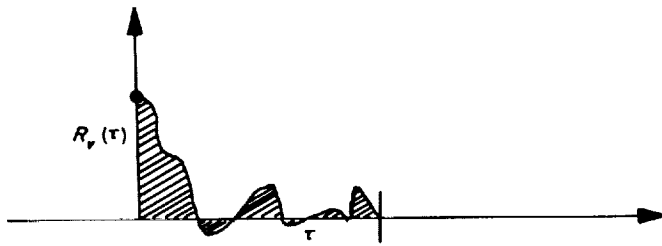$$\sigma_{wu}^2 = \frac{2}{T} \int_{0}^{T} \left(1 - \frac{\tau}{T}\right) R_u\left(\tau\right) d\tau.$$

Fig. 10. The correlation of the aperiodic part of $R_v(\tau)$

Then for $T > \alpha$, we have

$$\sigma_{wu}^2 = \frac{2}{T} \left[\beta + \varepsilon(T) - \delta(T) \right]$$

where we have substituted

$$\varepsilon(T) = \int_\alpha^T R(\tau)\, d\tau$$

$$\delta(T) = \frac{1}{T} \int_0^T \tau R_u(\tau)\, d\tau.$$

It is then clear that $|\varepsilon(T)| < \varepsilon$, and

$$|\delta(\tau)| \leq \int_0^\alpha \frac{\tau}{T} |R_u(\tau)|\, d\tau + \int_\alpha^T \frac{\tau}{T} |R_u(\tau)|\, d\tau$$

$$\leq \frac{\alpha B}{T} + \varepsilon.$$

This along with σ_{wp}^2 produces the result

$$\left| \sigma_w^2 - \frac{2\beta}{T} \right| < \frac{4\varepsilon}{T} + \frac{1}{T^2}(K + 2\alpha B).$$

Thus, if we choose $\varepsilon << \beta/2$, then for $T >> K + 2\alpha B$ and $T > \alpha$, we see that

$$\sigma_w^2 \approx \frac{2\beta}{T}$$

which provides the result that the output squared-mean-to-variance ratio λ_0 is asymptotically equal to

$$\lambda_0 \approx \frac{T\mu^2}{2\beta}$$

under the conditions that $\varepsilon << \beta/2$, $T > \alpha$, and $T >> K + 2\alpha B$.

From this it is evident that λ_0 *ultimately increases linearly in T for all input processes and inversely as the area under the autocorrelation function.* The computation of λ_0 then becomes one of estimating μ^2 and β.

4. The Squaring Loop Technique for Binary PSK Synchronization, J. J. Stiffler

a. Introduction. A previous summary (Ref. 16) investigated the problem of optimum power allocation between the data and a synchronization signal in a two-channel binary PSK communiction system. For such a system, however, another method of synchronization is immediately suggested by the following observation: The data signal is of the form

$$s(t) = 2^{1/2} A m(t) \sin(\omega_0 t + \phi) \qquad m(t) = \pm 1,$$

$$vT_b < t < (v+1) T_b, \qquad v = 0, 1, \cdots$$

Squaring this signal yields

$$s^2(t) = A^2 \{ 1 - \cos 2(\omega_0 t + \phi) \}$$

which contains an unmodulated frequency component at twice the subcarrier frequency. Tracking this squared data signal with a phase-locked loop and halving the output frequency should therefore establish subcarrier synchronization. This method is analyzed in part *b* of this article. In part *c* it is shown that a similar technique may be used to establish bit synchronization and in part *d* the analysis is extended to include the case of simultaneous bit and subcarrier synchronization. Comparisons are made with the two-channel system and error probability curves associated with the two methods when both subcarrier and bit synchronization must be obtained are included. It is concluded that, in any realistic situation, the squaring technique provides superior performance. In fact, even if both methods were available and could be used complementally, in almost all cases the optimum allocation would leave no power in the synchronization channel.

b. Subcarrier synchronization. The adaptation of the phase locked loop shown in Fig. 11 presumably will enable subcarrier synchronization to be established. It is necessary to determine the behavior of the phase error θ_e. It will be observed that this process leaves a 180-deg phase ambiguity since both $(+ \sin(\omega t + \phi))^2$ and $(- \sin(\omega t + \phi))^2$ yield the same signal. However, this ambiguity may often be resolved in some other way and is not necessarily objectionable. It is assumed that such is the case here.
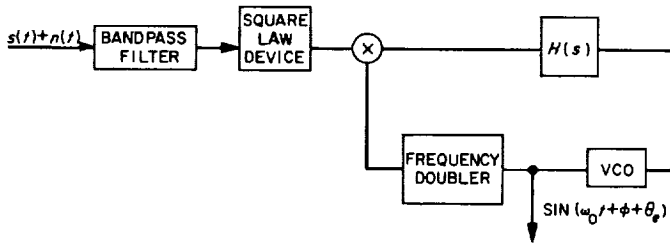
Fig. 11. The squaring loop

The input to the bandpass is $2^{1/2}A\,m\,(t)\sin(\omega_0 t+\phi)+n(t)$ where $n(t)$ represents white noise and $m(t)=\pm 1$. The bandwidth W of the filter is assumed to be wide enough so that the signal output is essentially unchanged. A partial compensation for the effect on the signal of the filter may be made by reducing the effective signal amplitude by an amount r (Ref. 17) where r^2 is the ratio of the filtered signal power to that of the unfiltered signal. Since the modulating signal is assumed to be a random sequence of pulses of unit amplitude and time duration T_b,

$$S_m(f)=\frac{T_b\sin^2\omega T_b/2}{(\omega T_b/2)^2}$$

then

$$r^2=\frac{\displaystyle\int_{-2\pi W}^{2\pi W}\frac{\sin^2(\omega T_b/2)}{(\omega T_b/2)}\,d\omega}{\displaystyle\int_{-\infty}^{\infty}\frac{\sin^2(\omega T_b/2)^2}{(\omega T_b/2)^2}\,d\omega}. \tag{1}$$

The bandwidth W will presumably be such that, at the least, all signal energy out to the first null in the spectrum is passed. Then $W\gtrsim 1/T_b$ and $r^2\gtrsim 0.91$. Then the output of the filter is assumed to be

$$2^{1/2}rA m(t)\sin(\omega_0 t+\phi)+n_1(t)=rS(t)+n_1(t) \tag{2}$$

where $n_1(t)$ is noise with the spectrum

$$S_{n_1}(f)=N_0 \qquad \left|\frac{\omega_0}{2\pi}-W\right|\leq f\leq\left|\frac{\omega_0}{2\pi}+W\right|$$

$$=0 \qquad\qquad \text{elsewhere} \tag{3}$$

The square-law device then produces the signal

$$y(t)=r^2A^2(1-\cos 2(\omega_0 t+\phi))$$

$$+2(2)^{1/2}rA m(t)n_1(t)\sin(\omega_0 t+\phi)+n_1^2(t)$$

$$=[rs(t)+n_1(t)]^2. \tag{4}$$

It is necessary to determine the signal and noise power in a narrow region centered about $\omega=2\omega_0$. To accomplish this, consider the autocorrelation function (Ref. 18)

$$R_y(\tau)=E\{(rs_1+n_1)^2(rs_2+n_2)^2\}$$

$$=r^4E[s_1^2 s_2^2]+4r^2[s_1 s_2]E[n_1 n_2]+r^2E[s_1^2]E[n_2^2]$$

$$+r^2E[s_2^2]E[n_1^2]+E[n_1^2 n_2^2] \tag{5}$$

where $s_i=s(t_i)$, $n_i=n_1(t_i)$, $\tau=t_1-t_2$, and use is made of the fact that the expected value of the noise is zero and that the signal and noise are independent. Proceeding to evaluate the terms:

$$E[s_1^2 s_2^2]=A^4\left(1+\frac{1}{2}\cos 2\omega_0\tau\right)$$

$$E[s_1 s_2]=R_s(\tau)$$

$$E[n_1 n_2]=R_{n_1}(\tau)$$

$$E[s_1^2]E[n_2^2]=E[s_2^2]E[n_1^2]=R_s(0)R_{n_1}(0)$$

$$E[n_1^2 n_2^2]=E^2[n_1 n_2]+2E(n_1^2)E(n_2^2)$$

$$=R_{n_1}^2(\tau)+2R_{n_1}^2(0) \tag{6}$$

The last equation makes use of the fact that the noise is Gaussian.

Transforming the autocorrelation functions to get the spectral densities of the various terms yields:

$$S_{s\times s}(\omega)=\frac{r^4A^4}{4}[\delta(f-2f_0)+\delta(f+2f_0)]$$

and, at $\omega=2\omega_0$,

$$S_{s\times n}(2\omega_0)=4r^2\int_{-\infty}^{\infty}R_s(\tau)R_{n_1}(\tau)e^{-j2\omega_0\tau}\,d\tau$$

$$=4r^2\int_{-\infty}^{\infty}S_{n_1}(f)S_s(2f_0-f)\,df$$

$$=4r^2N_0\int_{f_0-W}^{f_0+W}\frac{A^2T_b}{2}\left\{\frac{\sin^2\left[(\omega_0-\omega)\dfrac{T_b}{2}\right]}{\left[(\omega_0-\omega)\dfrac{T_b}{2}\right]^2}\right.$$

$$\left.+\frac{\sin^2\left[(3\omega_0-\omega)\dfrac{T_b}{2}\right]}{\left[(3\omega_0-\omega)\dfrac{T_b}{2}\right]^2}\frac{d\omega}{2\pi}\right\} \tag{7}$$

$$\approx 2r^2A^2N_0$$

$$S_{n \times n}(2\omega_0) = \int_{-\infty}^{\infty} R_{n_1}^2(\tau) e^{-j2\omega_0\tau} d\tau$$

$$= \int_{-\infty}^{\infty} S_{n_1}(f) S_{n_1}(2f_0 - f) df$$

$$= N_0^2 \int_{f_0-W}^{f_0+W} df = 2WN_0^2.$$

The notation $S_{s \times s}(\omega)$ refers to that portion of the power spectrum due to the product of the signal with itself, $S_{s \times n}(2\omega_0)$ that portion at $\omega = 2\omega_0$ due to the product of the signal with the noise, and $S_{n \times n}(2\omega_0)$ that due to the noise with itself. Since only the first term represents useful information the second two terms act as noise. At the frequency $\omega = 2\omega_0$ then, the signal power is approximately $r^4A^4/2$ while the single-sided noise spectral density is

$$4r^2A^2N_0 + 4WN_0^2. \tag{8}$$

Before proceeding it is important to make several observations: First, the noise at the input to the phase-locked loop is not white, and second, it is not Gaussian. The important factor here is that presumably the loop bandwidth B_L is very small compared to the bandwidth W of the bandpass filter. The noise input to the loop thus looks nearly flat over the frequency range of the loop, and it is not unreasonable to assume that it is white. In addition, it has been shown (Ref. 19) that when the input to the network of Fig. 12 is a modulated sinusoid plus white Gaussian noise, the statistics of the output approach Gaussian statistics as $W_1/W_2 \to \infty$. Thus, under the assumption that $W >> B_L$ an approximate solution to the problem at hand may be obtained by assuming that the input noise to the phase-locked loop is white and Gaussian with the normalized power spectral density

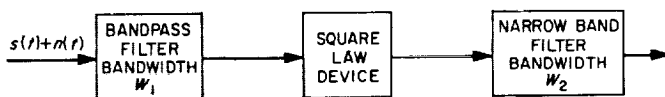$$8\left(\frac{N_0}{r^2A^2} + \frac{WN_0^2}{r^4A^4}\right).$$

Fig. 12. A squaring network

If the bandwidth of the loop is B_L' then the variance of the double frequency output error is approximately

$$\sigma_{2\theta}^2 = 8B_L'\left(\frac{N_0}{r^2A^2} + \frac{WN_0^2}{r^4A^4}\right). \tag{9}$$

Since the frequency of the phase jitter after the frequency doubler is twice that before the doubler, let $B_L' = 2B_L$ where B_L is the loop bandwidth that would be necessary to track the frequency before doubling.

$$\sigma_\theta^2 = \frac{1}{4}\sigma_{2\theta}^2 = 4B_L\left\{\frac{N_0}{r^2A^2} + \frac{WN_0^2}{r^4A^4}\right\}$$

$$= \frac{4}{Q}\left\{\frac{1}{r^2R_T^2} + \frac{k}{r^4R_T^4}\right\} \tag{10}$$

where

$$W = \frac{k}{T_b}, \quad R_T^2 = \frac{P_T}{N_0}T_b, \quad P_T = A^2, \text{ and } Q = \frac{1}{T_bB_L}.$$

It was shown in Ref. 16 that in a two-channel system (the second channel being used to provide subcarrier synchronization for the first), the signal-to-noise ratio $R_s^2 = P_s/N_0B_L$ in the sync channel should satisfy the equation

$$R_s^2 \approx R^2\left[1 + \left(1 + \frac{Q}{R^2}\right)^{\frac{1}{2}}\right] \tag{11}$$

where $R^2 = ST_b/N_0$, the signal-to-noise ratio in the data channel, and $Q = 1/T_bB_L$ the reciprocal of the product of the bit time and the loop bandwidth $\hat{f}$, as before. This was the condition for the optimum division of power between the two channels when the total power was constrained. The parameter R_s could be increased only by decreasing the value of R.

The question arises as to whether the squaring loop provides an accurate enough estimation of the phase. Should a second channel be used, in addition, to provide an improved estimate? Since the second channel power would be available at the expense of the first channel, this question may be rephrased as follows: Is the phase estimate provided by the squaring loop as accurate as that which would be provided by a second channel under an optimum division of power? If the answer is yes, then clearly the power in the data channel should not be decreased to increase the synchronization accuracy since

the latter is already at least as precise as it would optimally be in the absence of a squaring loop Then a second loop is not needed if

$$\frac{1}{\sigma_\theta^2} \geq \left[1 + \left(1 + \frac{Q}{R_T^2} \right)^{\frac{1}{2}} \right] R_T^2$$

or if

$$1 \geq \frac{4}{Q} \left\{ \frac{1}{r^2 R_T^2} + \frac{k}{r^4 R_T^4} \right\} R_T^2 \left[1 + \left(1 + \frac{Q}{R_T^2} \right)^{\frac{1}{2}} \right]$$

where, again, $R_T^2 = A^2 T_b / N_0$, the data signal-to-noise ratio, and $W = k/T_b$; thus, if

$$Q \geq 8 \left(\frac{1}{r^2} + \frac{k}{r^4 R_T^2} \right) \left[1 + 2 \left(\frac{1}{r^2 R_T^2} + \frac{k}{r^4 R_T^4} \right) \right] \approx 8. \quad (12)$$

Since in most practical situations Q will be at least 100, this situation seems to be easily satisfied. It has been assumed that $k \geq 1$ (and that Q be reasonably large) in order that the approximations made in determining this result be valid. Actually it is seen that k can be much larger than one and the inequality (12) still remains satisfied. The larger k is the more accurate the approximation that $W >> B_L$ and that the only effect of filtering the signal before squaring it is that the effective signal energy is reduced by the factor r^2. Since increasing k narrows the squared signal spectrum and decreasing k decreases the noise power, there is presumably some optimum value of k. Because of the nonlinearities involved, it is difficult to determine theoretically the optimum value of k, although it could be observed experimentally. It is believed that this optimum value will not be much greater than one or two since 91% of the signal energy is included in the bandpass filter of bandwidth $2/T_b$ and 95% in the filter of bandwidth $4/T_b$. Increasing the bandwidth further only adds more noise without significantly changing the signal spectrum.

c. Bit synchronization. Bit synchronization can also be obtained with a squaring loop. If, for example, the signal $(2A)^{\frac{1}{2}} m(t) \sin \omega_b t$ is used to represent a bit with $m(t) = \pm 1$, $\nu T_b < t < (\nu + 1) T_b$, and $\omega_b = \pi/T_b$, then the squared signal is $A^2 (1 - \cos 2\omega_b t)$. The double frequency component may be tracked with a phase-locked loop as before to establish the phase and thereby the bit timing of the received signal. Again, there is a 180-deg phase ambiguity which may or may not be objectionable, depending upon the system. Since the signal consists of
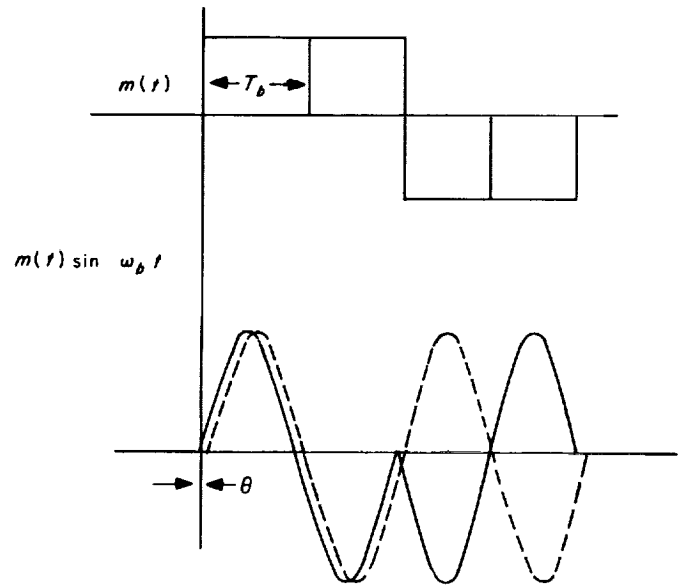

Fig. 13. The effect of mis-synchronization

a sequence of pulses of the form $m(t) \sin \omega_b t$ rather than square pulses, it is necessary to reconsider the effect of mis-synchronization. This may be done by referring to Fig. 13.

It can be seen that if the phase error is θ radians and two successive pulses have the same sign, the signal amplitude is effectively reduced by the factor $\cos \theta$ as was the case with subcarrier mis-synchronization. However, if the two successive pulses are of different sign, then this amplitude reduction is slightly less due to the end effect. The exact amplitude reduction factor is easily determined to be

$$\cos \theta + \frac{2}{\pi} (\sin \theta - \theta \cos \theta)$$

which, for all values of θ of interest, is greater than $\cos \theta$. For small θ, moreover, the difference is slight and for practical values of σ_θ, the difference may be neglected. Consequently, this is the same situation as that encountered in the problem of subcarrier synchronization and, if two channels are to be used, the allocation of power between them is as it was in the case of subcarrier synchronization. Hence, for reasonable values of Q, the squaring loop alone suffices and a second channel need not be used. The only difference between obtaining bit synchronization and obtaining subcarrier synchronization with a squaring loop is that the bandpass filter of the latter becomes a low-pass filter in the former and the

bandwidth will be approximately half that in the previous situation to pass the same fraction of the signal energy.

It is interesting to notice that the error due to mis-synchronization is less using sinusoidal pulses rather than square pulses. It is doubtful that this advantage is sig-

nificant, however, particularly in view of the fact that the generation of sinusoidal pulses may be less efficient.

d. Combined subcarrier and bit synchronization. In any real situation it is generally necessary to obtain *both* subcarrier and bit synchronization. While it is perhaps possible to use the former to increase the accuracy of the latter, it is dubious that this is practical in most situations since the subcarrier frequency will generally be much greater than the bit frequency.

The error probability could be expressed as a function of both the bit and subcarrier phase errors (Ref. 16), and the allocation of a power among the three channels optimized. Clearly, the mathematics would get rather cumbersome. A solution which is probably as satisfactory and certainly considerably easier follows from the observation that for large values of Q, the error probability has not changed much due to the lack of perfect synchronization. Thus, the power may be optimized, as in Ref. 16, between the data and the subcarrier sync channels, assuming perfect bit synchronization and between the data and bit sync channels, assuming perfect subcarrier synchronization. The second order effects, for large Q, are presumably negligible. Using this method of power division, the error probability was determined as a function of the total signal-energy-per-bit-to-noise-spectral-density ratio

$$\frac{P_T T_b}{N_0} = R^2 + \frac{R_{sb}^2}{Q} + \frac{R_{sc}^2}{Q} = \frac{(S + P_{sb} + P_{sc}) T_b}{N_0}$$

and of the value of Q, when both subcarrier and bit synchronization must be determined. (The bandwidth B_L is assumed to be the same for both loops, and the subscripts b and c refer to the bit and subcarrier sync channels, respectively.) The expression for the bit-error probability is, in this case,

$$P_E = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\theta_b) p(\theta_c) P_E(\theta_b, \theta_c) d\theta_b d\theta_c$$

$$= \frac{1}{2^{1/2} \pi^{3/2}} \int_0^{\infty} e^{-x^2/2} \int_0^{\infty} e^{-y^2/2} \int_{2^{1/2}R \cos \sigma_c y}^{\infty} e^{-z^2/2} dz \, dy \, dx$$

$$+ \frac{1}{2^{1/2} \pi^{3/2}} \int_0^{\infty} e^{-x^2/2} \int_0^{\infty} e^{-y^2/2} \int_{2^{1/2}R[1 - (\sigma_b x/\pi)] \cos \sigma_c y}^{\infty} \quad (13)$$

The term σ_b^2 is the variance of the bit phase estimate, and σ_c^2 the variance of the subcarrier phase estimate. The results are plotted as a function of Q in Fig. 14.
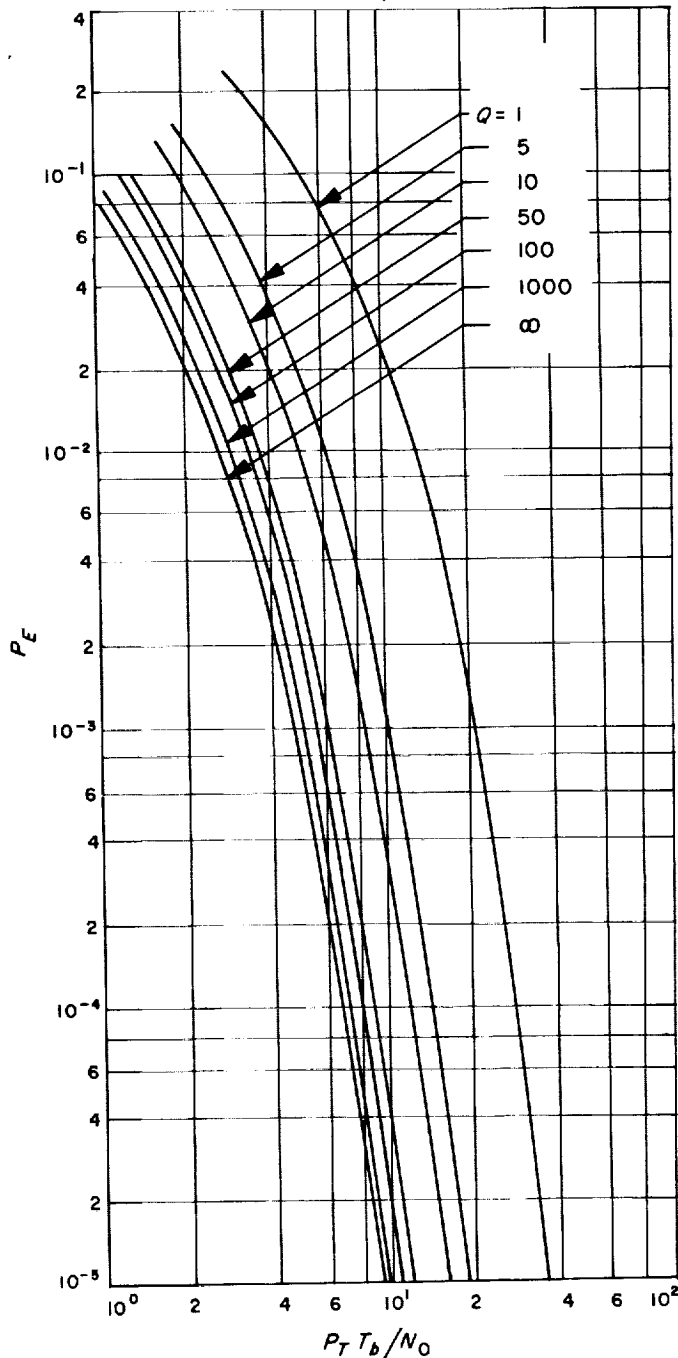

Fig. 14. Bit error probability using separate subcarrier and bit synchronization channels

If a squaring loop is used, then, since

$$(Am(t)\sin \omega_b t \sin \omega_c t)^2 = A^2\left[1 - \cos 2\omega_b t - \cos 2\omega_c t \right.$$

$$\left. + \frac{1}{2}\cos 2(\omega_c - \omega_b)t + \frac{1}{2}\cos 2(\omega_c + \omega_b)t \right] \quad (14)$$

both the subcarrier and bit synchronization can be obtained sufficiently accurately without the use of auxiliary channels so long as the unequality (12) is satisfied for both Q_b and Q_c, the Q's corresponding to the bit and subcarrier channels, respectively. Since the signal is now $m(t) \sin \omega_b t \sin \omega_c t$ with $\omega_b = \pi/T_b$, the spectrum has the form shown in Fig. 15 and the bandpass filter preceding the square-law device should have a bandwidth $2W$ of at least $3/T_b = 3\omega_b/\pi$. The bit-error probability was also determined, when the squaring loop is used, as a function of the ratio $P_T T_b/N_0$ and of Q, where as before, it is assumed that $Q_b = Q_c = Q$. The estimate of the phase error variance of Eq. (10) was used, and the error probability determined as before. The subcarrier and bit phase error variance σ_c^2 and σ_b^2 are now determined from Eq. (10). The bandwidth $2W$ is selected to be $3/T_b$. The results are plotted in Fig. 16.

e. *Conclusion.* It is readily seen from Figs. 14 and 16 that, if possible, it is preferable to use a squaring loop to obtain bit and subcarrier synchronization. In addition, it may be concluded from the analysis that it would not be advantageous to use any of the available power to transmit an unmodulated subcarrier in most situations if a squaring loop is available. This supports a conclu-
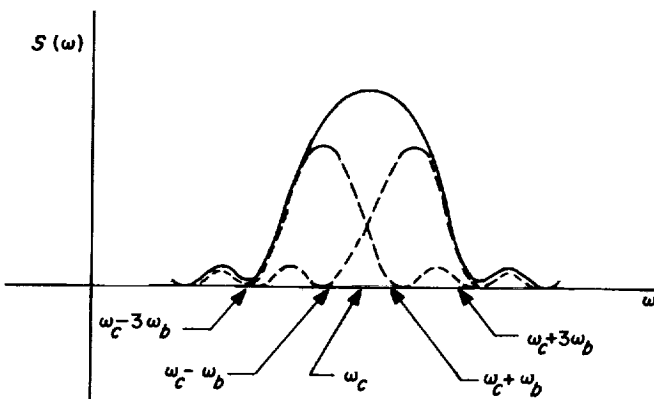

Fig. 15. The power spectrum of the signal
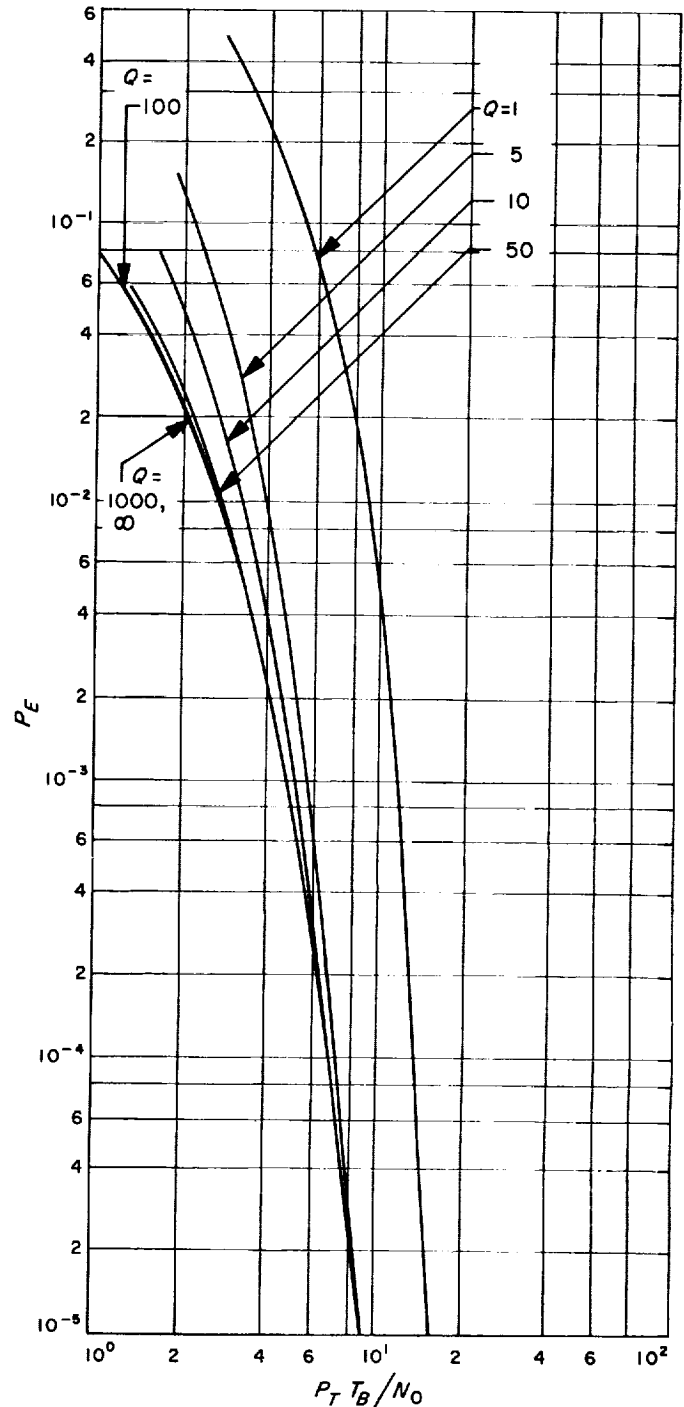$m(t) \sin \omega_b t \sin \omega_c t$

Fig. 16. Bit error probability using a squaring loop
for bit and subcarrier synchronization

sion arrived at by Van Trees (Ref. 17) in which the allocation of power between the data and a sync signal was considered, a squaring loop being used in conjunction with the former. The information from the two channels was to be combined optimally to establish subcarrier synchronization. Although the analysis of such a system

becomes rather difficult, Van Trees concluded that, in all situations he was able to investigate, all the power should be relegated to the data. The analysis presented here shows that this will always be true so long as the ratio of the bit rate to the loop bandwidth satisfies the inequality (12).

5. The Probability Density Function for Phase-Difference Estimates Obtained from the Envelopes of the Sum and Difference of Two Sine Waves, W. B. Kendall

The problem of measuring the phase difference between two low-level sine waves frequently arises in RF systems, e.g., in interferometric angle measuring systems. It has been shown (Ref. 20) that, in the presence of additive Gaussian noise, the maximum likelihood estimate of this phase difference is obtained by taking the difference of the estimates of the phases of the two signals with respect to an arbitrary reference signal. This is shown schematically in Fig. 17. The two amplifiers for this system must be phase matched, but their gains need not be the same. Both the sign and the magnitude of the phase difference are obtained by this system.

Occasionally phase-difference measurements between low-level, equi-amplitude sine waves must be made using amplifiers which are *not* phase matched, but which *are* gain matched. Then a possible estimation procedure is that diagrammed in Fig. 18. Note that this system obtains an estimate of only the *magnitude* of the phase difference. In what follows we will obtain an exact expression for the probability density function of this phase-difference-magnitude estimate.

If the amplifiers in Fig. 18 add wide-band Gaussian noise to their sine wave inputs, then the probability density functions of the outputs of the envelope detectors are the Rice distribution (Ref. 21):

$$p(z) = ze^{-(z^2+\alpha^2)/2} I_0(\alpha z)$$

where

Z = envelope detector output

$z = Z/(\psi_0)^{\frac{1}{2}}$

$\alpha = A/(\psi_0)^{\frac{1}{2}}$

$\psi_0 = N_0/T$

A = input sine wave peak amplitude

N_0 = one-sided noise spectral density referred to the amplifier input

T = detector integration time

The probability distribution function of the ratio of two independent Rice variates z_- and z_+ with parameters α_- and α_+, respectively, is (Ref. 22)

$$Pr\left\{\frac{z_-}{z_+} \leq r\right\} = Q\left(\frac{r\alpha_+}{(1+r^2)^{\frac{1}{2}}}, \frac{\alpha_-}{(1+r^2)^{\frac{1}{2}}}\right)$$
$$- \frac{1}{1+r^2}\exp\left[-\frac{(r\alpha_+)^2 + (\alpha_-)^2}{2(1+r^2)}\right] I_0\left(\frac{r\alpha_+\alpha_-}{1+r^2}\right)$$

$S_1(t) = a \sin(\omega t + \phi + y/2)$

$f_1(t) = G_1 S_1(t) + n_1(t)$

Phase Detector — θ_1

G_1

Amplifiers

$\sin \omega t$

G_2

Phase Detector — θ_2

$f_2(t) = G_2 S_2(t) + n_2(t)$

$S_2(t) = a \sin(\omega t + \phi - y/2)$

$n_i(t)$ = Gaussian-noise sample functions

$\theta_i = \tan^{-1}(X_i/Y_i)$

$X_i = \frac{2}{T}\int_0^T f_i(t) \sin \omega t\, dt$

$Y_i = \frac{2}{T}\int_0^T f_i(t) \cos \omega t\, dt$

y = phase-difference estimate
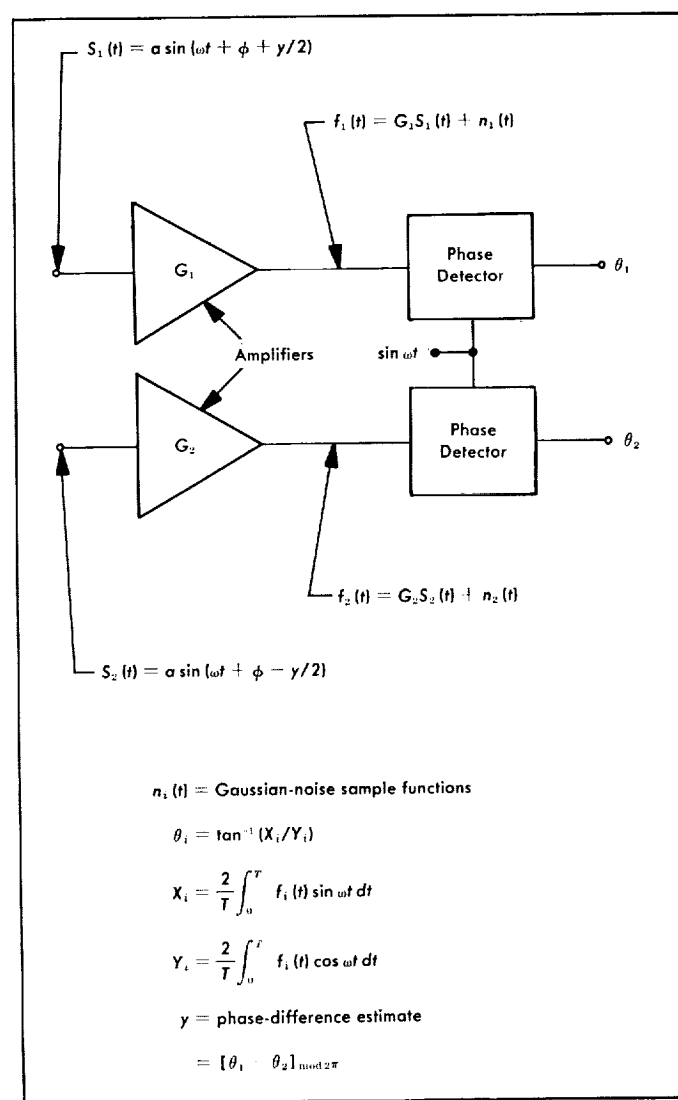
$= [\theta_1 - \theta_2]_{\text{mod}\,2\pi}$

Fig. 17. Maximum-likelihood phase-difference estimator

where $Q(a, b)$ is the Q-function (Ref. 23) defined by

$$Q(a, b) = \int_b^\infty x \exp\left[-\frac{x^2 + a^2}{2} \right] I_0(ax)\, dx.$$
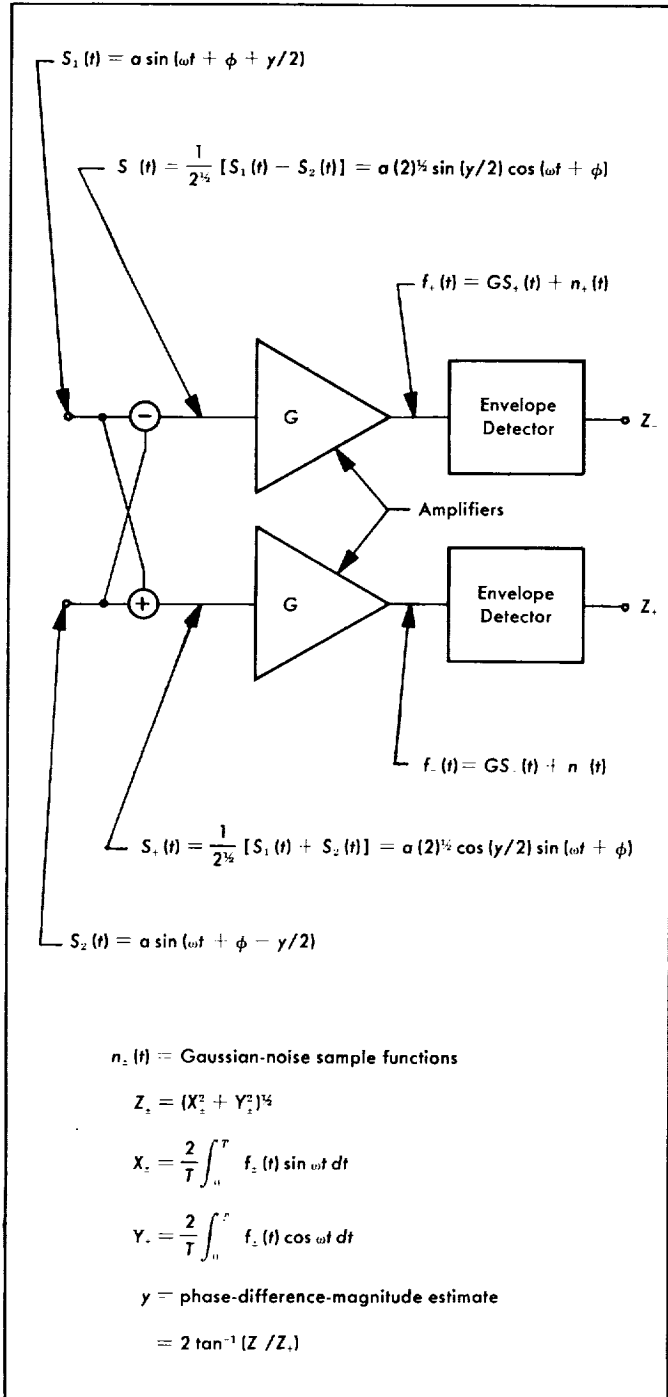

Fig. 18. Alternate phase-difference estimator which uses gain- rather than phase-matched amplifiers

Now, if the noise spectral density is the same for both amplifiers in Fig. 18, we have

$$Pr\left\{ \tan\left(\frac{y}{2}\right) = \frac{Z_-}{Z_+} \leq r \right\} = Pr\left\{ \frac{z_-}{z_+} \leq r \right\}$$

where now

$\alpha_- = a(2)^{1/2} \sin(y/2)\, (T/N_0)^{1/2} = 2(E/N_0)^{1/2} \sin(y/2)$

$\quad = 2(R)^{1/2} \sin(y/2)$

$\alpha_+ = 2(R)^{1/2} \cos(y/2)$

E = energy of *each* received signal during time T

$R = E/N_0$ = signal-to-noise ratio for each signal

Next, the relations

$$\frac{\partial Q(a, b)}{\partial a} = b \exp\left[-\frac{b^2 + a^2}{2} \right] I_1(ab)$$

and

$$\frac{\partial Q(a, b)}{\partial b} = -b \exp\left[-\frac{b^2 + a^2}{2} \right] I_0(ab)$$

can be used to differentiate

$$Pr\left\{ \frac{z_-}{z_+} \leq r \right\}$$

to obtain

$$p_{\tan (\hat{y}/2)}(r) = \frac{d}{dr} Pr\left\{ \frac{z_-}{z_+} \leq r \right\}$$

$$= \frac{2r}{(1 + r^2)^2} \exp\left[-\frac{(r\alpha_+)^2 + (\alpha_-)^2}{2(1 + r^2)} \right]$$

$$\times \left\{ \left[1 + \frac{(\alpha_+)^2 + (r\alpha_-)^2}{2(1 + r^2)} \right] I_0\left(\frac{r\alpha_+\alpha_-}{1 + r^2} \right) \right.$$

$$\left. + \left[\frac{r\alpha_+\alpha_-}{1 + r^2} \right] \qquad I_1\left(\frac{r\alpha_+\alpha_-}{1 + r^2} \right) \right\}.$$

Finally, the probability density function of the estimated phase-difference magnitude $\hat{y}$ given the true phase difference y can be found by substituting for α_- and α_+ in the relation

$$p_{\hat{y}}(\hat{y}|y) = \left[\{p_{\tan (\hat{y}/2)}(r)\} \times \left| \frac{d}{d\hat{y}} \tan(\hat{y}/2) \right| \right]_{r = \tan (\hat{y}/2)}$$

$$= \left[\frac{1 + r^2}{2}\, p_{\tan (\hat{y}/2)}(r) \right]_{r = \tan (\hat{y}/2)}$$

The result is

$$p_{\hat{y}}(\hat{y}|y) = \frac{1}{2} \sin \hat{y}\, e^{-R\,(1-\cos y \cos \hat{y})}$$

$$\times \{[1 + R\,(1 + \cos y \cos \hat{y})]\, I_0\,(R \sin y \sin \hat{y})$$

$$+ [R \sin y \sin \hat{y}] \qquad\qquad I_1\,(R \sin y \sin \hat{y})\}.$$

Note that this function is symmetric about the point $(y, \hat{y}) = (\pi/2, \pi/2)$, i.e.,

$$p_{\hat{y}}(\pi - \hat{y}|\pi - y) = p_{\hat{y}}(\hat{y}|y).$$

The exponential behavior of $p_{\hat{y}}$ can be made more evident by representing it in terms of the slowly varying functions

$$\mathscr{I}_0(x) = e^{-x} I_0(x) \sim \frac{1}{(2\pi x)^{\frac{1}{2}}}\left[1 + \frac{1}{8x} + \frac{9}{128 x^2} + \cdots \right]$$

and

$$\mathscr{I}_1(x) = e^{-x} I_1(x) \sim \frac{1}{(2\pi x)^{\frac{1}{2}}}\left[1 - \frac{3}{8x} - \frac{15}{128 x^2} - \cdots \right]$$

Use of these allows $p_{\hat{y}}$ to be written

$$p_{\hat{y}}(\hat{y}|y) = \frac{1}{2}\sin \hat{y}\, e^{-R\,[1-\cos(\hat{y}-y)]}$$

$$\times \{[1 + R\,(1 + \cos y \cos \hat{y})]\, \mathscr{I}_0\,(R \sin y \sin \hat{y})$$

$$+ [R \sin y \sin \hat{y}] \qquad\qquad \mathscr{I}_1\,(R \sin y \sin \hat{y})\}.$$

This is the probability density function we seek. It is shown in Fig. 19 for signal-to-noise ratio $R = 4$, 20, and 100. From this figure it is evident that $\hat{y}$ is a biased estimate of y except when $y = \pi/2 = 90$ deg.

When the true phase difference y is zero, $p_{\hat{y}}$ becomes

$$p_{\hat{y}}(\hat{y}|0) = \frac{1}{2}\,[1 + R\,(1 + \cos \hat{y})]\sin \hat{y}\, e^{-R\,(1-\cos \hat{y})}.$$

Then for the high signal-to-noise ratio case in which $R \gg 1$, we have, in the vicinity of the peak,

$$\cos \hat{y} \approx 1 - \hat{y}^2/2 \approx 1$$

and

$$\sin \hat{y} \approx \hat{y}$$

so that

$$p_{\hat{y}}(\hat{y}|0) \approx R\hat{y}\, e^{-R\hat{y}^2/2}$$

which is the Rayleigh distribution with mean $(\pi/2R)^{\frac{1}{2}}$ and mean-squared value $2/R$.

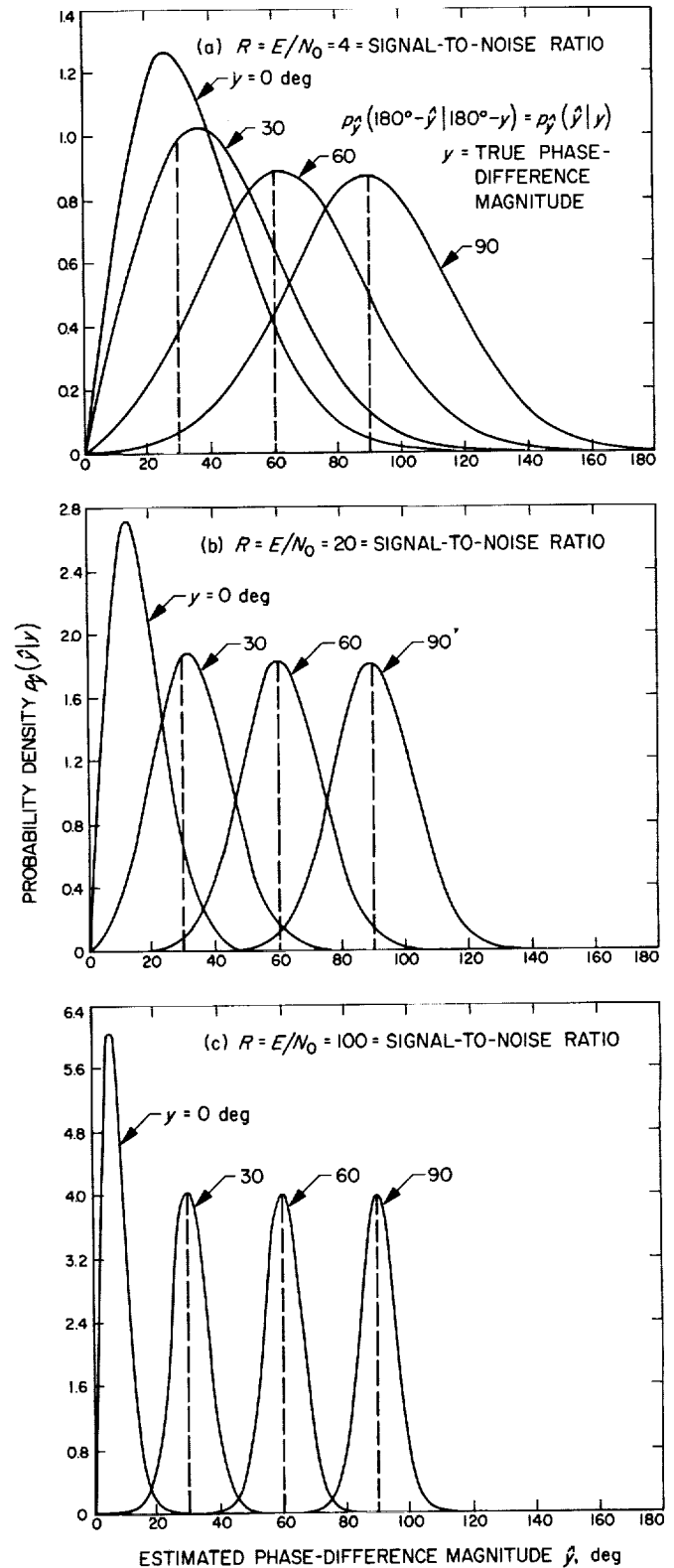
Fig. 19. Probability density function for the phase-difference-magnitude estimate obtained by the estimator of Fig. 18

For the high signal-to-signal ratio case in which $R \sin y \sin \hat{y} >> \frac{3}{8}$, the functions $\mathcal{J}_0$ and $\mathcal{J}_1$ can be represented by only the first terms of their asymptotic series, and $p_{\hat{y}}$ can be approximated by

$$p_{\hat{y}}(\hat{y}|y) \approx \left(\frac{\sin \hat{y}}{8\pi \sin y}\right)^{\frac{1}{2}} R\,[1 + \cos(\hat{y} - y)]\,e^{-R[1 - \cos(\hat{y} - y)]}.$$

Then, in the vicinity of the peak, we have

$$\cos(\hat{y} - y) \approx 1 - \frac{(\hat{y} - y)^2}{2} \approx 1;$$

and so for very large signal-to-noise ratios and $y \neq 0$, we get

$$p_{\hat{y}}(\hat{y}|y) \approx \frac{R}{(2\pi)^{\frac{1}{2}}} e^{-R(\hat{y} - y)^2/2},$$

which is the Gaussian distribution with variance $1/R$. This is the same as the distribution of the estimate of the maximum-likelihood estimator of Fig. 17 at high signal-to-noise-ratios. Thus, we see that the performance of this suboptimum estimator approaches that of the maximum likelihood estimator for high signal-to-noise ratio R as long as the true phase-difference magnitude is greater than about $2/R^{\frac{1}{2}}$ radians.

6. Predetection Filters in Radiometric Receivers,
R. C. Titsworth

a. Introduction. In a radiometric receiver, the presence and strength of a weak signal in noise is determined by estimating the power of signal-plus-noise and subtracting from this an estimate of the noise power. It is often convenient in planetary and lunar radar systems to perform this signal estimation by keying the transmitter at some slow periodic rate, so that the receiver observes equal bursts of signal-plus-noise and noise-only wave-forms per keying cycle. When the signal return is very narrow band, to reject noise it is desirable to use a very narrow-band IF filter. But because of the time-constants involved in narrow-band filtering, the IF waveform has its signal portion spilling over into the noise-only part, and this degrades the estimate. This article investigates this phenomenon.

b. The receiver. Fig. 20 shows the form of a planetary or lunar radar receiver. All the RF and IF filtering is represented by a bandpass filter. The detector is assumed to be a perfect square-law device. During the "S + N" part of the receiver cycle, the output of the squarer is integrated; during the "N" portion of the cycle, the negative of the squared output is integrated into the estimate, tending to subtract out the noise. At the end of the cycle the contents of the integrator are $S + v$, where S is proportional to the signal power, and v is a zero-mean noise term.

Only frequencies centered about the carrier are selected by the predetection filter; the square-law device output is then a term centered at zero-frequency and a high-frequency term. This latter term is rejected by the integrator, and can be omitted. The output behaves, then, as if the received wave were a low-frequency series, filtered by a low-pass, squared, and then integrated.

We shall adopt this equivalent model, shown in Fig. 21, for analysis here. We assume that the input $x(t)$ to the receiver is a square-wave $s(t)$ in stationary white Gaussian noise $n(t)$, and we normalize the amplitude of $s(t)$ to 1.

$$s(t) = \begin{cases} 1 & (0 \le t < 1) \\ 0 & (1 \le t < 2) \end{cases}$$
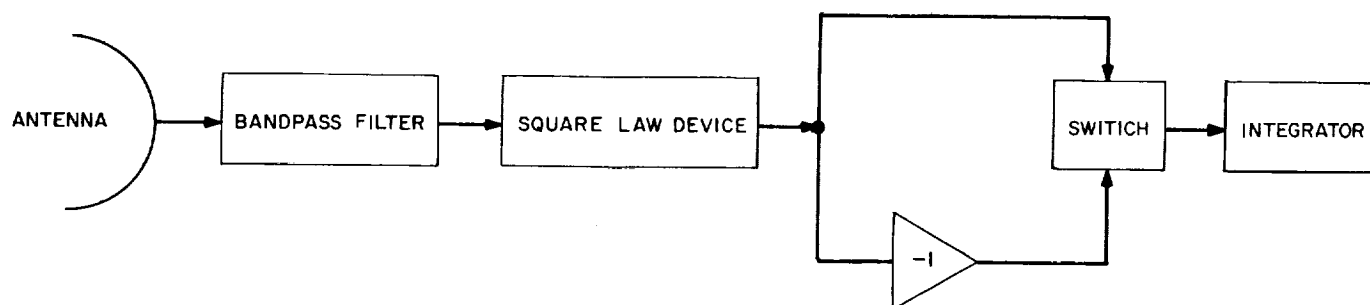
$$R_n(\tau) = \frac{1}{2} N_0 \delta(\tau).$$
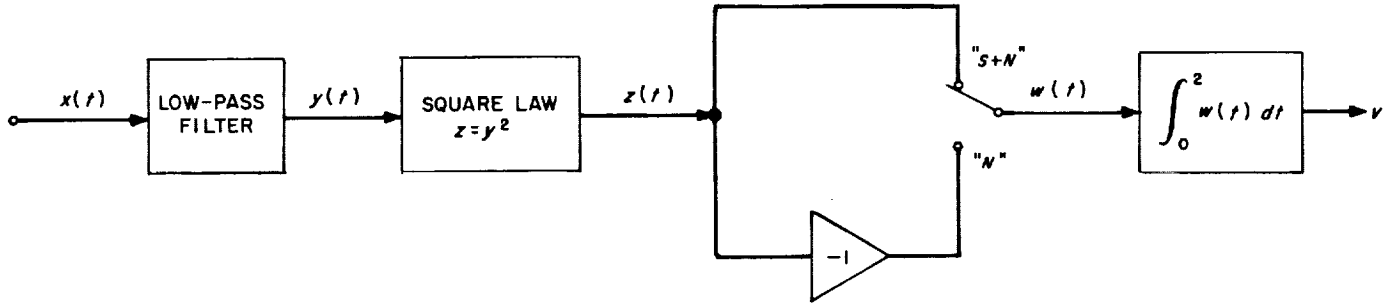

Fig. 20. Planetary or lunar receiver

Fig. 21. Model of receiver for analysis

We have also scaled time so that the keying cycle has period 2. N_0 is the (single-sided) noise spectral density.

The filtered version of $x(t)$ appears at the output of the low-pass filter as

$$y(t) = a(t) + N(t)$$

in which we represent by $a(t)$ and $N(t)$ the response of the filter to $s(t)$ and $n(t)$, respectively. Obviously, as indicated in Fig. 22,

$$a(t) = 1 - a(t+1) = 1 - a(t-1).$$

At the end of the keying cycle, the output v of the integrator is

$$v = \int_0^1 \{[a^2(t) - a^2(t-1)] + [N^2(t) - N^2(t-1)]$$

$$+ 2[a(t)N(t) - a(t-1)N(t-1)]\} dt.$$

The mean of v, call it μ, is then the estimate of the signal strength.
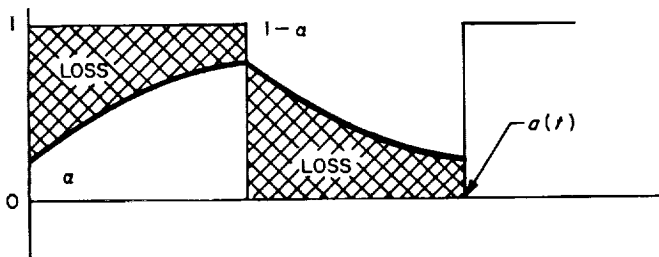
Fig. 22. Response of the filter to the keyed signal

This differs from the "true" value by an amount dependent on the filter characteristic; it is easy to see $\mu \approx 1$ if a wide-band filter is used and $\mu \approx 0$ when the bandwidth is very narrow.

Presumably, even though μ differs from unity, it differs by a fixed amount for a fixed filter. As a performance criterion then, we may compare the squared-mean-to-variance (SMV) ratios as the filter bandwidth is varied.

The computation of the variance of v, denote it σ^2, is straightforward, and

$$\sigma^2 = 4 \int_0^1 \int_{-u}^{1-u} \left\{ R^2(\tau) - \frac{1}{2} R^2(\tau+1) - \frac{1}{2} R^2(\tau-1) \right.$$

$$+ [a(\tau+u)a(u) + a(\tau+u-1)a(u-1)]R(\tau)$$

$$- a(\tau+u)a(u-1)R(\tau+1)$$

$$\left. - a(\tau+u-1)a(u)R(\tau-1) \right\} d\tau\, du.$$

Use here has been made of the fact that $N(t)$ is a zero-mean Gaussian process, and $R(\tau)$ denotes its autocorrelation function.

We have tacitly assumed that the filter $H(s)$, whose impulse response we shall denote by $h(t)$, has unit gain at dc. We shall also assume the gain at any other frequency is less than one. Hence, $|H(j\omega)| \leq H(0) = 1$.

It follows that

$$R(\tau) = \frac{1}{2} N_0 \int_0^\infty h(u)h(u+\tau)\, du$$

and if B denotes the (single-sided) noise-bandwidth of $H(s)$, $R(0) = N_0 B$, so

$$B = \frac{1}{2} \int_0^\infty h^2(u)\, du.$$

As an example, let us assume the filter is a one-section RC type with

$$h(t) = 4B\, e^{-4Bt}.$$

It is easy to verify that B is the noise bandwidth of the filter, and that

$$R(\tau) = N_0 B\, e^{-4B|\tau|}.$$

The response $a(t-1)$ is given by (Fig. 22)

$$a(t-1) = \frac{e^{-4Bt}}{1 + e^{-4B}}$$

$$a(t) = \frac{1 + e^{-4B} - e^{-4Bt}}{1 + e^{-4B}} \qquad (0 \le t < 1)$$

so that the mean integrator output is

$$\mu = 1 - \frac{1}{2B}\left(\frac{1 - e^{-4B}}{1 + e^{-4B}}\right).$$

We shall also assume for simplicity that N_0 is sufficiently large that the dominant terms in σ^2 are those involving N_0^2:

$$\sigma^2 \approx 4 \int_0^1 \int_{-u}^{1-u} \left[R^2(\tau) - \frac{1}{2} R^2(\tau - 1) \right.$$
$$\left. - \frac{1}{2} R^2(\tau + 1) \right] d\tau\, du.$$

Using the expression above

$$\sigma^2 \approx N_0^2 \left\{ B\left[1 - \frac{1}{8B}(1 - e^{-8B}) \right] - \frac{1}{16}(1 - e^{-8B})^2 \right\}$$

$$\approx N_0^2 \left\{ B - \frac{1}{16}(1 - e^{-8B})(3 - e^{-8B}) \right\}.$$

Then the SMV ratio λ is μ^2/σ^2. Let $\gamma = e^{-4B}$. Then

$$\lambda = \frac{\left[1 - \frac{1}{2B}\left(\frac{1-\gamma}{1+\gamma}\right) \right]^2}{N_0^2(1+\gamma)^2\left\{ B - \frac{1}{16}(1-\gamma^2)(3-\gamma^2) \right\}}$$

$$= \frac{[2B(1+\gamma) - 1 + \gamma]^2}{4B^2 N_0^2(1+\gamma)^2\left\{ B - \frac{1}{16}(1-\gamma^2)(3-\gamma^2) \right\}}.$$

A plot of $(N_0^2\lambda)$ appears in Fig. 23. Note the very interesting behavior of this curve. For very small predetection
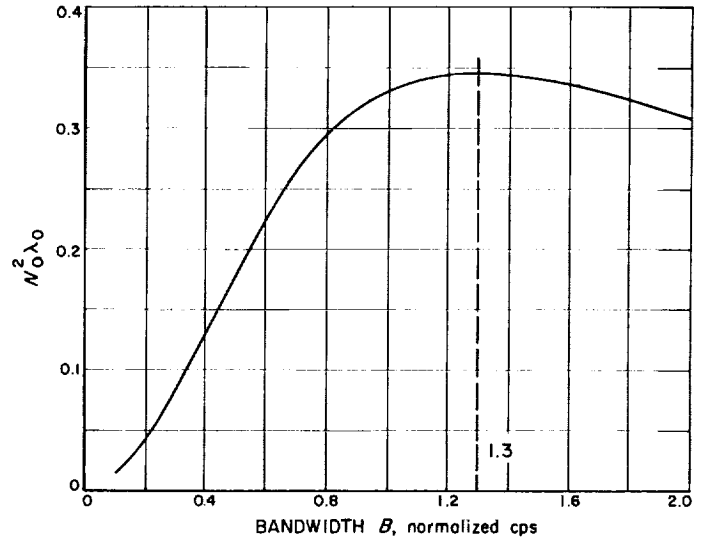
Fig. 23. The output SMV ratio $N_0^2\lambda_0$ of a radiometric receiver

bandwidth B, the signal spillover is overpowering the reduced noise power. On the other hand, at large bandwidths, the spillover is negligible, but the noise output is squared in the detector, and this causes great degradation at higher values of B. The best value of B, as shown in Fig. 23, is approximately $B = 1.3$. (Recall that time has been scaled so that the switching period is *two*.)

c. *Non-square law consideration.* Suppose, for a moment that we consider that the square-law device was not needed. In this case (Fig. 24)

$$v_1 = \int_0^1 [a(t) - a(t-1)]\, dt$$

$$+ \int_0^1 [N(t) - N(t-1)]\, dt.$$

The mean of v_1 is the same as before,
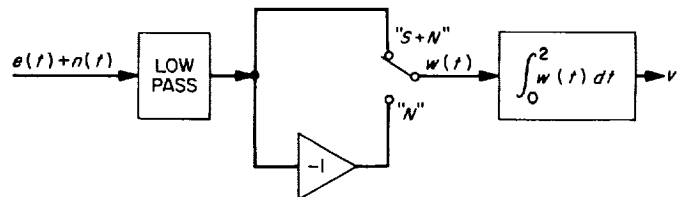
$$\mu = Ev_1 = 1 - 2\int_0^1 a(t-1)\, dt.$$

Fig. 24. A non-square-law idealization

But the variance of v_1 is

$$\sigma_1^2 = 2 \int_0^1 \int_{-u}^{1-u} \left[R(\tau) - \frac{1}{2} R(\tau - 1) \right. $$
$$\left. - \frac{1}{2} R(\tau + 1) \right] du\, d\tau.$$

Except for a constant scale factor, this resembles the first terms in σ^2 with $R(\tau)$ replacing $R^2(\tau)$.

Using the previous example for comparison

$$\sigma_1^2 = N_0 \left[1 - \frac{1}{4B} (1 - e^{-4B}) - \frac{1}{8B} (1 - e^{-4B})^2 \right].$$

The SMV ratio at the output is then

$$\lambda_1 =$$
$$\frac{[2B(1 + e^{-4B}) - 1 + e^{-4B}]^2}{4B^2(1 + e^{-4B})^2 N_0 \left[1 - \frac{1}{4B}(1 - e^{-4B}) - \frac{1}{8B}(1 - e^{-4B})^2 \right]}.$$

This takes the shape shown in Fig. 25. Note that λ_1 is monotone increasing in B; this indicates that if a square-law device were not needed, a wide predetection bandwidth is desirable. This is a drastically different result than obtained in the square-law case.

BANDWIDTH B, normalized cps
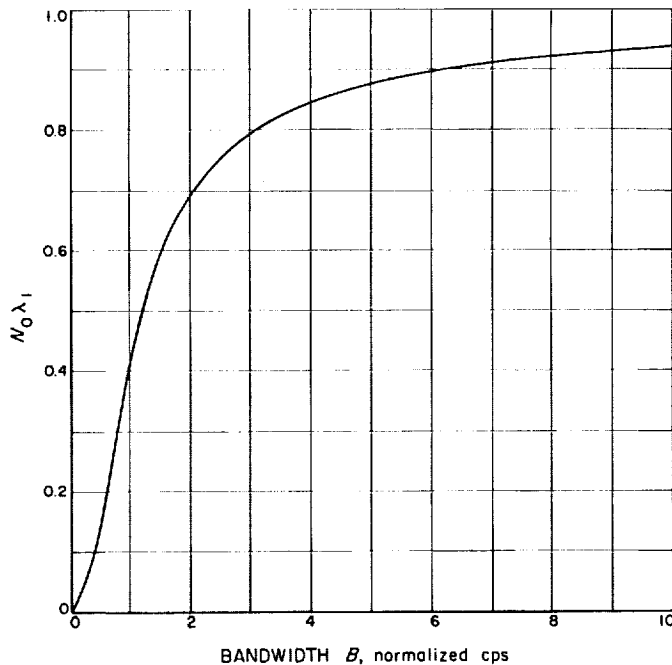
Fig. 25. The output SMV ratio $N_0\lambda_1$ of a linear radiometric receiver

7. Design of Varactor Harmonic Generators without Idlers, G. Thompson

a. Introduction. Varactors are used presently as harmonic generators in communication systems. For design purposes, the theoretical analysis of varactor harmonic generators assumes that the charge density under reverse bias is a step-function at the diode junction (Refs. 24 and 25). Such an abrupt-junction approach requires the use of idlers for all harmonic generators except the frequency doubler and the divide-by-two. With idlers the varactor harmonic generator provides high-efficiency frequency conversion. This fact has caused much emphasis to be placed upon the use of varactors at high frequencies where power amplification becomes difficult.

However, for low-frequency harmonic generation (100 kc to 300 Mc), high efficiency is not the primary concern of circuit design, and varactors without idlers will multiply and divide with sufficient output power to make subsequent amplification possible. The savings in number of components and complexity increase as the order of multiplication increases.

The purpose of this report is to present the initial progress made in determining proper criteria for the design of varactor harmonic generators without idlers. A departure from abrupt-junction theory has been made so that harmonic generation without idlers is theoretically possible.

As a first step, the frequency doubler is analyzed and the results are compared with the predictions for abrupt-junction theory. This comparison is valuable as a clue to the validity of the analysis since both abrupt-junction theory and the analysis to follow do not require an idler to multiply by two.

b. Equations of motion. The model of the varactor used in this analysis is shown in Fig. 26. The validity of this model is especially questionable at low frequencies because of the presence of leakage current across the junction. This current becomes most significant at low reverse bias voltages and small forward voltages. However, laboratory tests of Pacific Semiconductors, Inc.,
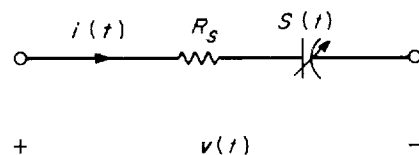

Fig. 26. The varactor model

(PSI) high-Q varicaps at 100 kc indicated that this leakage current, although appreciable, was still within tolerable limits. For this reason and for simplification of an already complicated problem, the model of Fig. 26 has been assumed to be valid.

Since it is convenient to think in terms of elastance, the voltage across the diode is expressed as

$$v(t) = \int S(t) i(t) dt + R_s i(t) \qquad (1)$$

where $v(t)$, $i(t)$, and R_s are defined in Fig. 26 and $S(t)$ is the elastance of the diode as a function of time. This equation is nonlinear since $S(t)$ depends on $i(t)$ in some complicated manner.

Following the Fourier analysis of harmonic generators in Ref. 24, we write:

$$v(t) = \sum_{k=-\infty}^{\infty} V_k e^{jk\omega_0 t} \qquad (2)$$

$$i(t) = \sum_{k=-\infty}^{\infty} I_k e^{jk\omega_0 t} \qquad (3)$$

and

$$S(t) = \sum_{k=-\infty}^{\infty} S_k e^{jk\omega_0 t} \qquad (4)$$

where ω_0 is the fundamental frequency and V_k, I_k, and S_k are the Fourier coefficients of $v(t)$, $i(t)$, and $S(t)$, respectively. Using Eq. (1) the Fourier coefficients of $v(t)$ are

$$V_k = \langle v(t) e^{-jk\omega_0 t} \rangle$$

$$V_k = R_s \langle i(t) e^{-jk\omega_0 t} \rangle + \frac{1}{jk\omega_0} \langle S(t) i(t) e^{-jk\omega_0 t} \rangle$$

$$V_k = R_s I_k + \frac{1}{jk\omega_0} \sum_{k=-\infty}^{\infty} I_l S_{k-l} \qquad (5)$$

where the brackets denote time average.

If $m\omega_0$ is the input frequency and $n\omega_0$ is the output frequency, then Eq. (5) for V_m and V_n will give steady-state equations of motion for a varactor harmonic generator. The condition for operation without idlers is that all I_l are assumed equal to zero except I_m and I_n.

c. Elastance–voltage characteristic. The relationship between the Fourier coefficients of elastance and the

current or charge can be derived from an elastance–voltage characteristic like that of Fig. 27. This plot shows the variation of diode elastance as a function of reverse bias voltage (taken to be positive). The curve was obtained by measuring the variation of capacitance with bias voltage of seven PSI high-Q varicaps on a Boonton 74C-S8 capacitance bridge. The test frequency was 100 kc and the test voltage was 50 mv rms. The nominal capacitance of the varicaps at -4 v ranged from 10 to 47 pf. The characteristic for each varicap was normalized to the value of capacitance at -4 v, C_r, and the seven curves were averaged to give Fig. 27.

If the curve is assumed to have the functional relationship

$$S \sim (v + k)^{1/2} \qquad (6)$$

then abrupt-junction theory applies and the Fourier coefficient, S_k, will be proportional to Q_k, the Fourier coefficient of charge. Idlers would then be necessary for harmonic generators with $m/n \neq 2$, $1/2$.

Since harmonic generators without idlers can be constructed with $m/n \neq 2$, $1/2$, then a higher degree of non-linearity than that of Eq. (6) is present in Fig. 27. A first approximation to the elastance–voltage characteristic was chosen to be

$$v = a_0 + a_2 S^2 + a_3 S^3. \qquad (7)$$

A least-squares fit to Fig. 27 yielded

$$a_0 = -0.79872165$$
$$a_2 = 6.1078186$$
$$a_3 = -0.021513835$$

To find S as a function of the charge, q, we note that

$$dv = S\,dq.$$

Differentiating Eq. (7) and substituting in Eq. (8) we can solve for dq such that

$$dq = (2a_2 + 3a_3 S)\, dS. \qquad (9)$$

Integrating both sides of Eq. (9) and taking the constant of integration to be the average charge, Q_0, we have

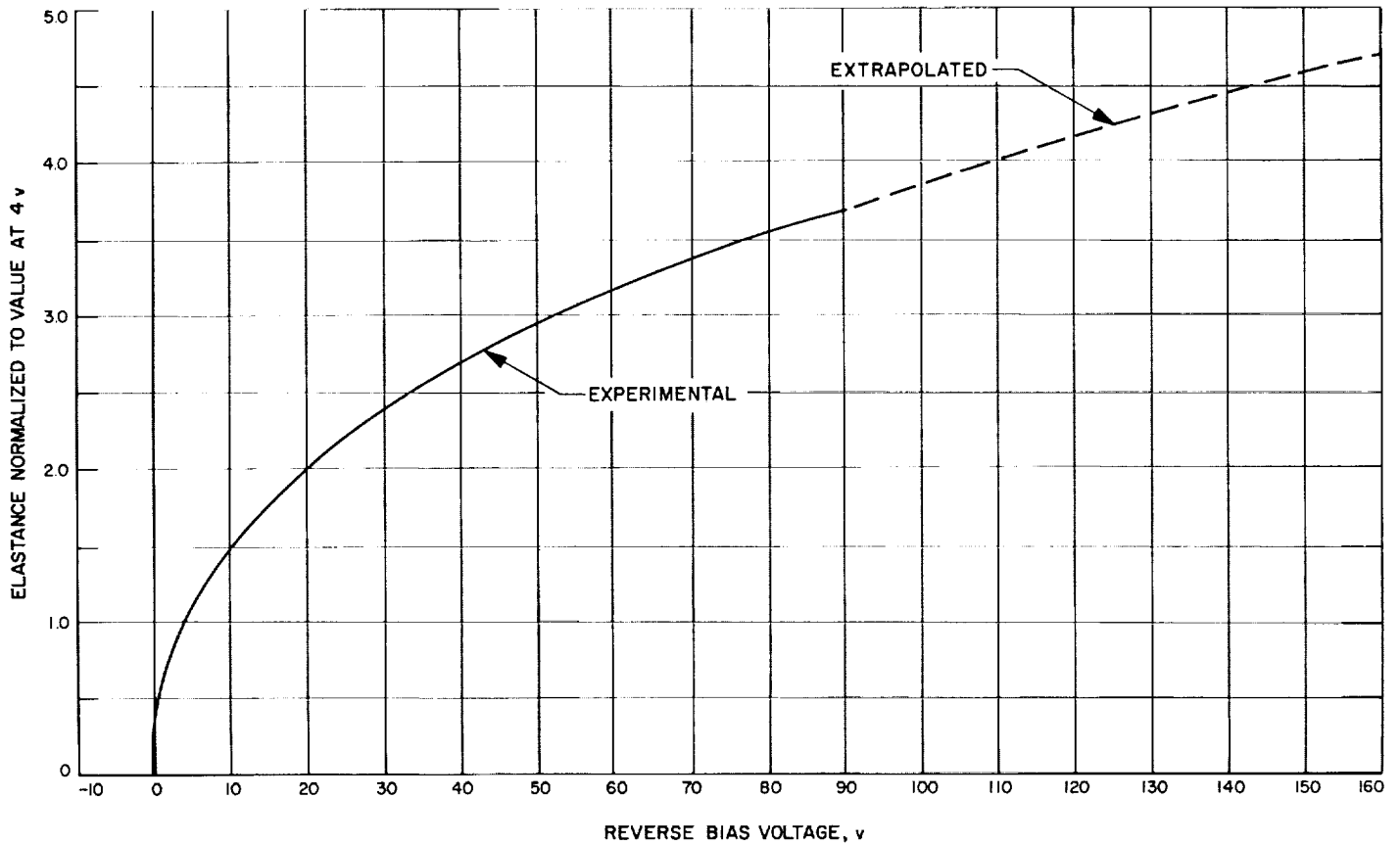$$q + Q_0 = 2a_2 S + \frac{3a_3}{2} S^2. \qquad (10)$$

Fig. 27. Elastance–voltage characteristic

After letting

$$A = 2a_2$$

and

$$B = 3a_3,$$

we solve Eq. (10) for S

$$S = \frac{-A \pm [A^2 + 2B(q + Q_0)]^{\frac{1}{2}}}{B}. \tag{11}$$

The choice of plus or minus sign is determined by making the value of the average elastance, S_0, positive over the range of variation of q.

We now assume that the input and output current coefficients are real and positive. The only justification for this assumption is that it simplifies the analysis. This assumption necessarily specifies the charge coefficients to be purely imaginary so that the ac variation of charge is

$$q(t) = -j|Q_m|(e^{jm\omega_0 t} - e^{-jm\omega_0 t}) - j|Q_n|(e^{jn\omega_0 t} - e^{-jn\omega_0 t}) \tag{12}$$

or

$$q(t) = 2|Q_m| \sin m\omega_0 t + 2|Q_n| \sin n\omega_0 t. \tag{13}$$

Therefore,

$$i(t) = 2m\omega_0 |Q_m| \cos m\omega_0 t + 2n\omega_0 |Q_n| \cos n\omega_0 t \tag{14}$$

so that

$$I_m = m\omega_0 |Q_m| \tag{15}$$

and

$$I_n = n\omega_0 |Q_n|. \tag{16}$$

If we substitute $q(t)$ from Eq. (13) into Eq. (11) and make the following identifications:

$$D = A^2 + 2BQ_0 \tag{17}$$

$$P_0 = \frac{D^{\frac{1}{2}}}{B} \tag{18}$$

$$P_m = \frac{4B |Q_m|}{D} \tag{19}$$

$$P_n = \frac{4B |Q_n|}{D}, \tag{20}$$

the resulting expression for S is

$$S(t) = -\frac{A}{B} \pm P_0 (1 + P_m \sin m\omega_0 t + P_n \sin n\omega_0 t)^{\frac{1}{2}}. \tag{21}$$

The Fourier coefficients of $S(t)$ can be obtained from Eq. (21) in the following manner. Expand the square root in a binomial series of the form

$$(1 + x)^{\frac{1}{2}} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{3x^2}{48} - \frac{15x^4}{384} + \cdots . \tag{22}$$

(In this analysis the series was terminated with the fourth power of x.) After using appropriate trigonometric identities, we can separate $S(t)$ into its frequency components.

d. The doubler equations. The analysis of the frequency doubler is made with $m = 1$ and $n = 2$ in Eq. (5). Including the assumption that I_1 and I_2 are real, we have

$$V_1 = \left(R_s + \frac{S_0}{j\omega_0} + \frac{S_2}{j\omega_0}\right) I_1 + \left(\frac{S_{-1}}{j\omega_0} + \frac{S_3}{j\omega_0}\right) I_2 \tag{23}$$

$$V_2 = \left(\frac{S_1}{j2\omega_0} + \frac{S_3}{j2\omega_0}\right) I_1 + \left(R_s + \frac{S_0}{j2\omega_0} + \frac{S_4}{j2\omega_0}\right) I_2. \tag{24}$$

Since $S(t)$ is real, $S_{-1} = S_1^*$. (The asterisk denotes complex conjugate.)

For greater generality we normalize the right-hand sides of these equations by dividing the elastance coefficients by S_r, the value of elastance at -4 v, and by multiplying the currents by S_r. We also normalize the frequency by defining

$$\omega_s = \frac{S_r}{R_s}. \tag{25}$$

Eqs. (23) and (24) then become

$$V_1 = R_{sn} \left[\left(1 + \frac{S_{0n}}{j\omega_{0n}} + \frac{S_{2n}}{j\omega_{0n}}\right) I_{1n} + \left(\frac{S_{-1n}}{j\omega_{0n}} + \frac{S_{3n}}{j\omega_{0n}}\right) I_{2n} \right] \tag{26}$$

$$V_2 = R_{sn} \left[\left(\frac{S_{1n}}{j2\omega_{0n}} + \frac{S_{3n}}{j2\omega_{0n}}\right) I_{1n} + \left(1 + \frac{S_{0n}}{j2\omega_{0n}} + \frac{S_{4n}}{j2\omega_{0n}}\right) I_{2n} \right] \tag{27}$$

where

$$R_{sn} = \frac{R_s}{S_r} = \frac{1}{\omega_s}. \tag{28}$$

The design parameters of interest are the load impedance, Z_2, the input "impedance," Z_1, the input and output powers, the efficiency, the bias voltage, and the breakdown limit. The load impedance is found by letting

$$V_2 = -Z_{2n} I_{2n}$$

and solving Eq. (27) for Z_{2n}. If

$$Z_{2n} = R_{2n} + j X_{2n} \tag{30}$$

then

$$R_{2n} = -R_{sn} \left[1 + \frac{S_{4ni}}{2\omega_{0n}} + \frac{(S_{1n} + S_3 n)_i}{2\omega_{0n}} \frac{I_{1n}}{I_{2n}} \right] \tag{31}$$

and

$$X_{2n} = \frac{R_{sn}}{2\omega_{0n}} S_{eon}. \tag{32}$$

S_{eon} is the normalized effective output elastance of the doubler and is

$$S_{eon} = \left[S_{0n} + S_{4n} + (S_{1n} + S_{3n}) \frac{I_{1n}}{I_{2n}} \right]_r. \tag{33}$$

The subscripts r and i indicate real and imaginary parts, respectively.

The input "impedance" is not really an impedance because the doubler is nonlinear and the ratio of V_1/I_1 depends on the drive level. Nevertheless, Z_1 is a useful design parameter. If

$$Z_{1n} = R_{1n} - j X_{1n} \tag{34}$$

then

$$R_{1n} = R_{sn} \left[1 + \frac{S_{2ni}}{\omega_{0n}} + \frac{(S_{-1n} + S_{3n})_i}{\omega_{0n}} \frac{I_{2n}}{I_{1n}} \right] \quad (35)$$

and

$$X_{1n} = \frac{R_{sn}}{\omega_{0n}} S_{ein}. \quad (36)$$

S_{ein} is the normalized effective input elastance of the doubler and is

$$S_{ein} = \left[S_{0n} + S_{2nr} + (S_{3n} + S_{-1n})_r \frac{I_{2n}}{I_{1n}} \right]. \quad (37)$$

The input power is then

$$P_1 = |I_1|^2 R_1 \quad (38)$$

and the output power is

$$P_2 = |I_2|^2 R_2. \quad (39)$$

The efficiency is

$$\epsilon = \frac{P_2}{P_1} = \left| \frac{I_2}{I_1} \right|^2 \frac{R_2}{R_1}. \quad (40)$$

The bias voltage, V_0, is obtained by taking the time average of Eq. (7). The result is

$$V_0 = a_0 + (a_2 + S_0 a_3)(S_0^2 + 2|S_1|^2 + 2|S_2|^2 + 2|S_3|^2 + \cdots). \quad (41)$$

The breakdown limit occurs at the value of reverse bias when the diode physically breaks down and begins
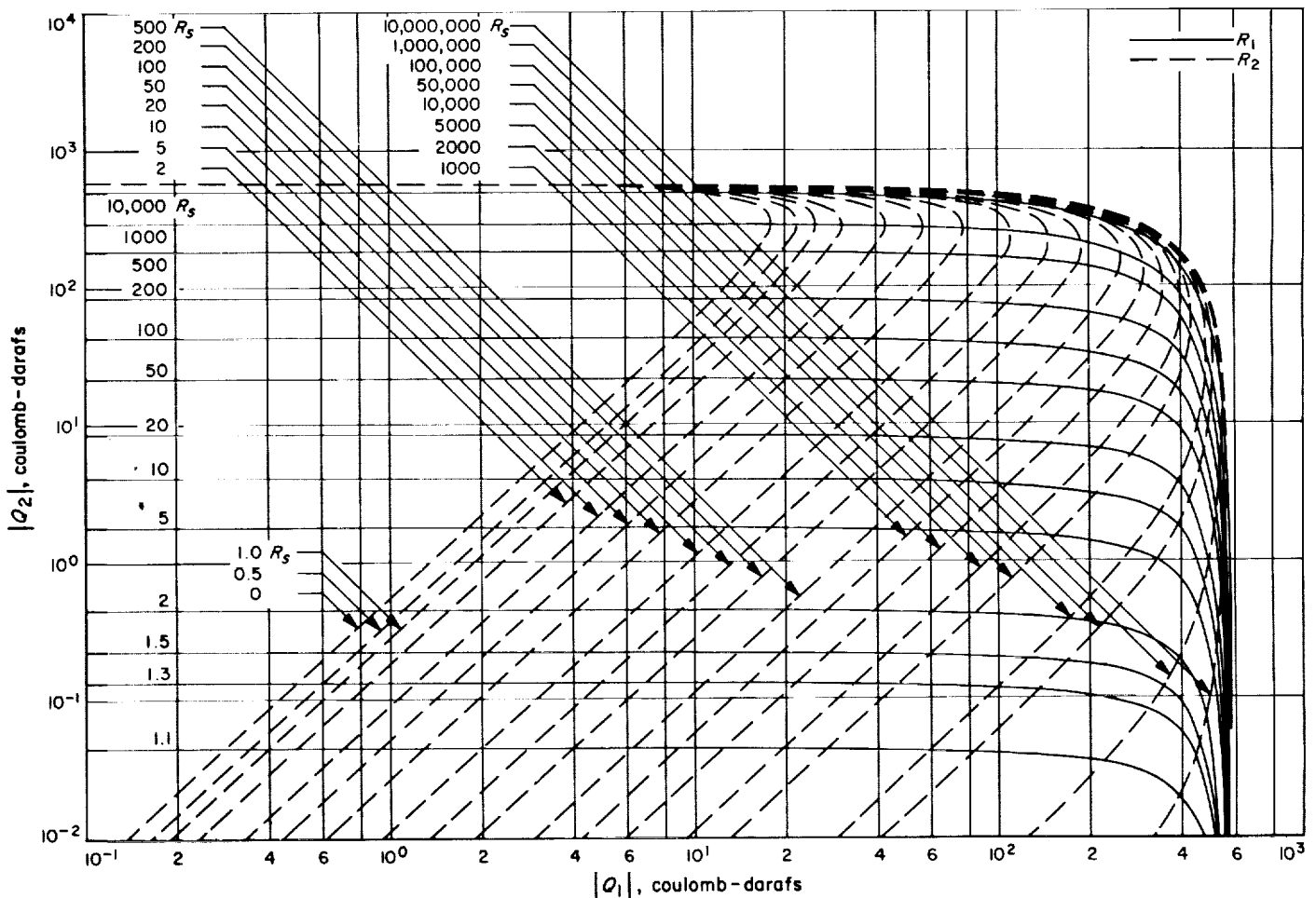
Fig. 28. Input and load resistances

to conduct. This value of voltage, V_b, should not be exceeded to maintain the validity of this analysis. We can remain within this limit as long as V_0 plus the positive peak of the ac variation of voltage is less than or equal to V_b.

The Fourier coefficients in Eqs. (23) and (24) are obtained from Eq. (21) with $m = 1$ and $n = 2$. Although the resulting expressions for these coefficients are complicated and generally unenlightening, they are included here for completeness.

$$S_0 = \frac{P_0}{2}\left[2 - \frac{(P_1 + P_2)}{8} - \frac{15}{512}(P_1^2 + P_2^2)^2 \right.$$

$$\left. - \frac{15}{256}P_1^2 P_2^2 \right] - \frac{A}{B} \tag{42}$$

$$S_1 = \frac{-P_0 P_1}{16}\left\{ P_2\left[1 + \frac{5}{32}(2P_1^2 + 3P_2^2) \right] \right.$$

$$\left. + j4\left[1 + \frac{3}{32}(P_1^2 + 2P_2^2) \right]\right\} \tag{43}$$

$$S_2 = \frac{P_0}{32}\left\{ P_1^2\left[1 + \frac{5}{32}(2P_1^2 + 3P_2^2) \right] \right.$$

$$\left. - j8P_2\left[1 + \frac{3}{32}(2P_1^2 + P_2^2) \right]\right\} \tag{44}$$

$$S_3 = \frac{P_0 P_1}{128}\left\{ 8P_2\left[1 + \frac{15}{32}(P_1^2 + P_2^2) \right] \right.$$

$$\left. + j\left[P_1^2 - 3P_2^2 \right]\right\} \tag{45}$$
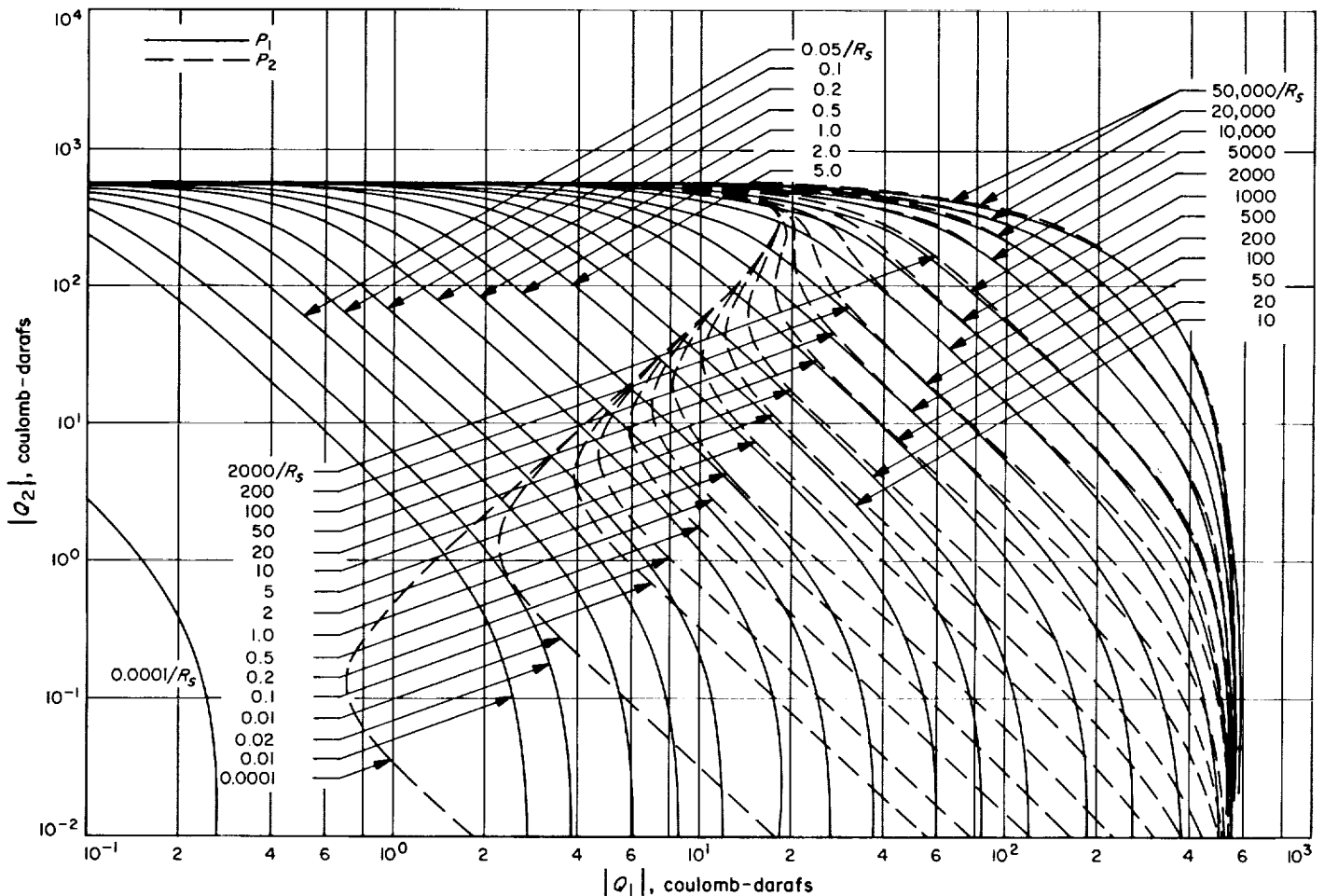
Fig. 29. Input and output powers

$$S_4 = \frac{P_0}{128} \left\{ 4 \left[P_2^2 - \frac{5}{64} \left(P_1^4 - 12P_1^2 P_2^2 - 4P_2^4 \right) \right] \right.$$

$$\left. + j3P_1^2 P_2 \right\} \tag{46}$$

$$S_5 = \frac{P_0 P_1 P_2}{512} \left[5 \left(P_2^2 - P_1^2 \right) + j12P_2 \right] \tag{47}$$

$$S_6 = \frac{P_0 P_2^2}{1024} \left[-15P_1^2 + j8P_2 \right] \tag{48}$$

$$S_7 = -\frac{5P_0 P_1 P_2^3}{512} \tag{49}$$

$$S_8 = -\frac{5P_0 P_2^4}{2048} \tag{50}$$

e. Results. The above design parameters were calculated as functions of $|Q_{1n}|$ and $|Q_{2n}|$ on the IBM 1620 computer. The results are shown in Figs. 28 through 32. The Q_1–Q_2 plane serves as a common area on which to plot the variation of the parameters. The exact values of Q_1 and Q_2 at a given operating point are really immaterial since the necessary design parameters are specified.

For a doubler S_{eon} and S_{ein} are essentially the same for the region of interest in the Q_1–Q_2 plane. Therefore, they are plotted on a single graph in Fig. 31. The plot of the breakdown voltage, V_b, in Fig. 30 requires some interpretation. The breakdown voltage of a particular diode must be determined. If the operating point of interest lies on the curve for this value of V_b, then the diode will be driven to its breakdown limit.

A comparison can be made between the results of this work and the results of abrupt-junction theory. Using
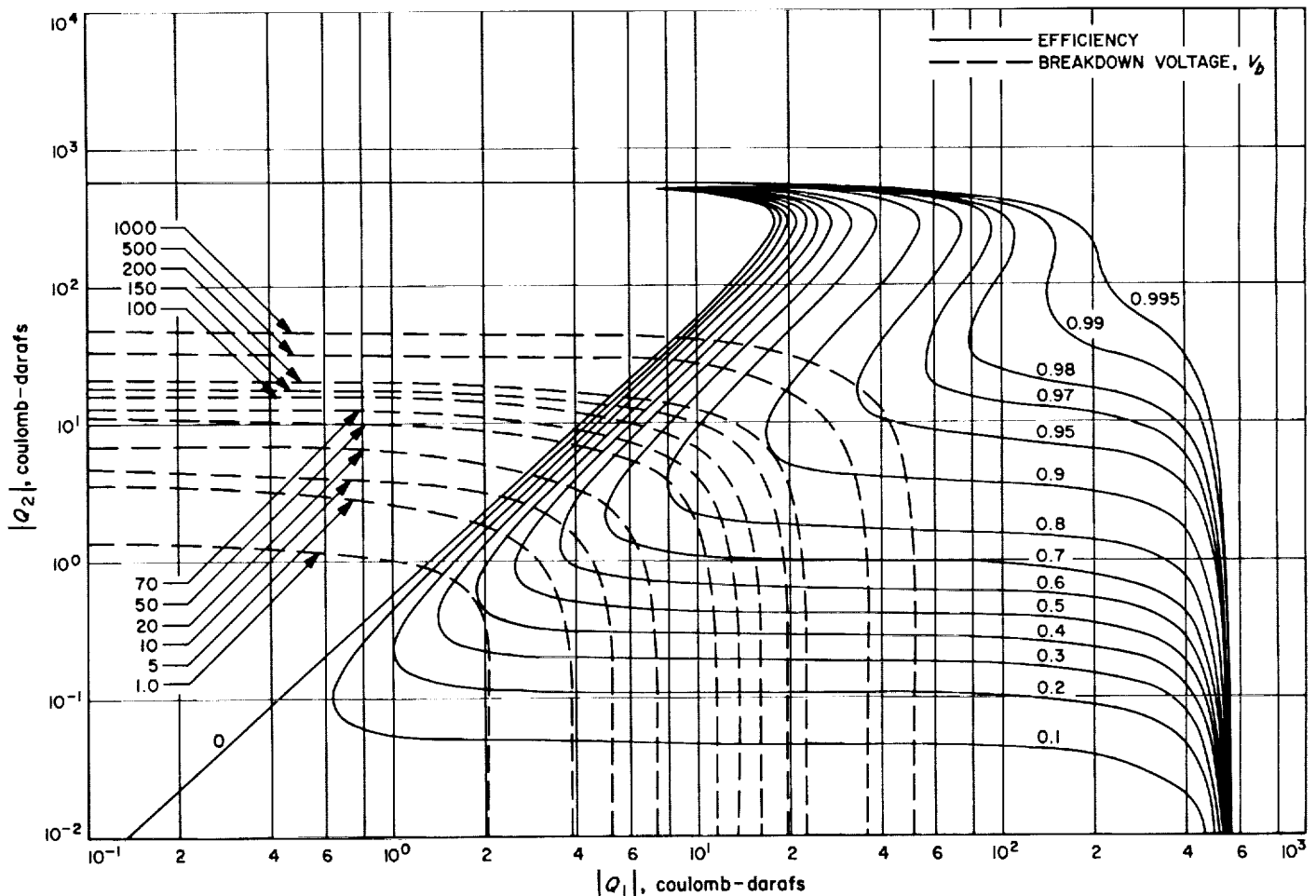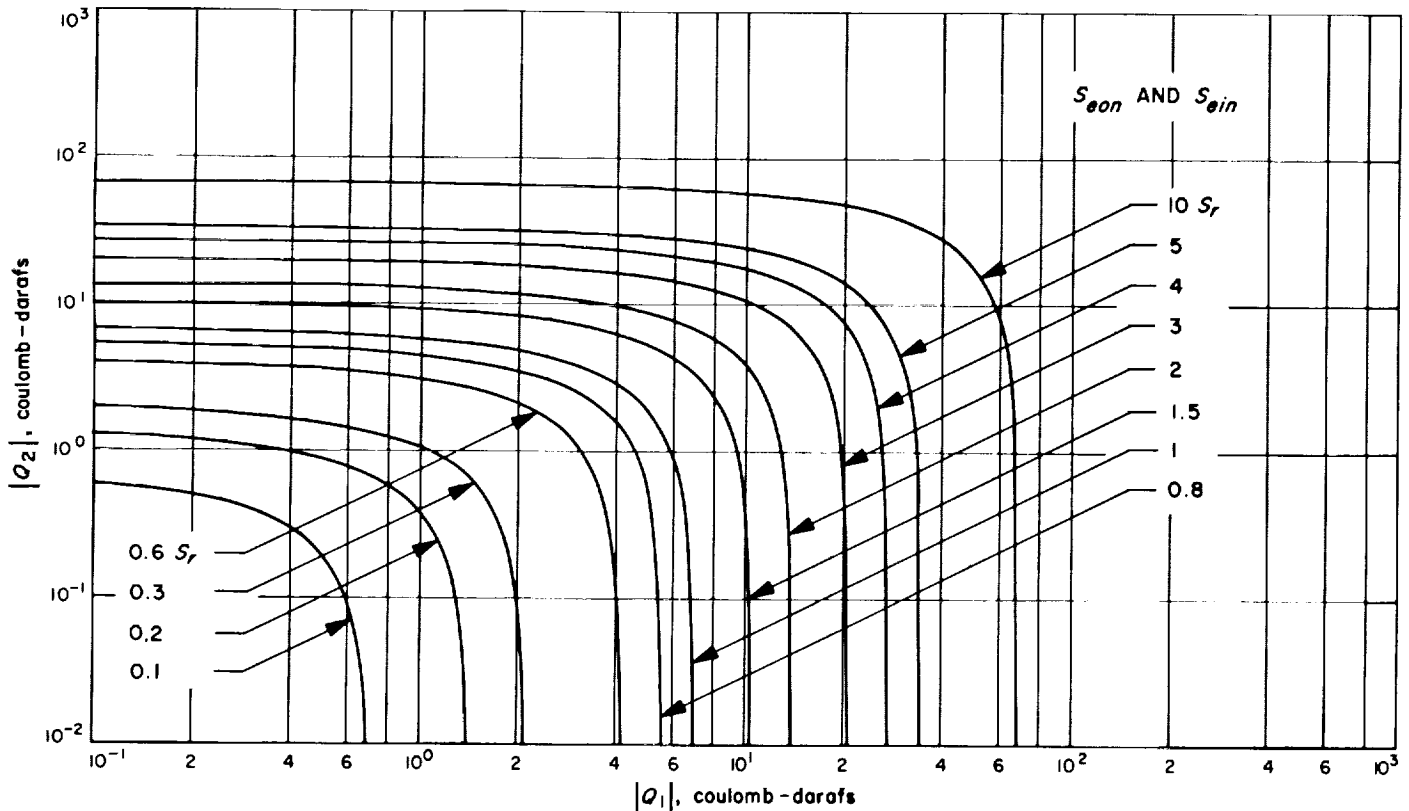

Fig. 30. Efficiency and breakdown limit

Fig. 31. Input and output effective elastances

the results given in Ref. 24, we plot the curves for the abrupt-junction doubler in Figs. 33 and 34. We have been careful to choose the same operating frequency in both analyses.

The ordinate in Figs. 33 and 34 is proportional to Q_2 and the abscissa is proportional to Q_1. From Ref. 24, the modulation ratio is defined as

$$m_k = \frac{|S_k|}{S_{max} - S_{min}} \propto Q_k \qquad (51)$$

where S_{max} and S_{min} are, respectively, the largest and smallest values on the elastance–voltage characteristic. Since these are constants for a particular diode, m_k is proportional to S_k. Because of the proportionality of S_k to Q_k in abrupt-junction theory, m_k is proportional to Q_k.

In effect the modulation ratio, m_k, is normalized with respect to the breakdown limit while Q_k is normalized with respect to S_r. Therefore, only one breakdown curve appears in Figs. 33 and 34, and there is a different curve for each value of V_b in Fig. 30. If we take S_{max} equal to

3.67S_r and S_{min} equal to zero in Fig. 27, then $V_b = 90$ v is the breakdown voltage in Fig. 30 which corresponds to that in Figs. 33 and 34.

We can now see that the two analyses are very similar. The maximum attainable efficiency is approximately 82% in both Figs. 30 and 34. The input power for this point is $2/R_s$ watts in Fig. 29 and $2.43/R_s$ watts in Fig. 33. The input and load resistances also have approximately the same functional form for both analyses. Further quantitative comparison will be made by experiment.

f. Conclusion. The preceding results give encouragement for further work into this approach to varactor multipliers. Experimental work is planned to illustrate the application of the above design procedure to a practical device and to compare experimental results with the two above analyses.

Since we have chosen a particular type of varactor to analyze, our results are, strictly speaking, limited. However, any diode that has an elastance–voltage characteristic that agrees with Fig. 27 to within 10% can be
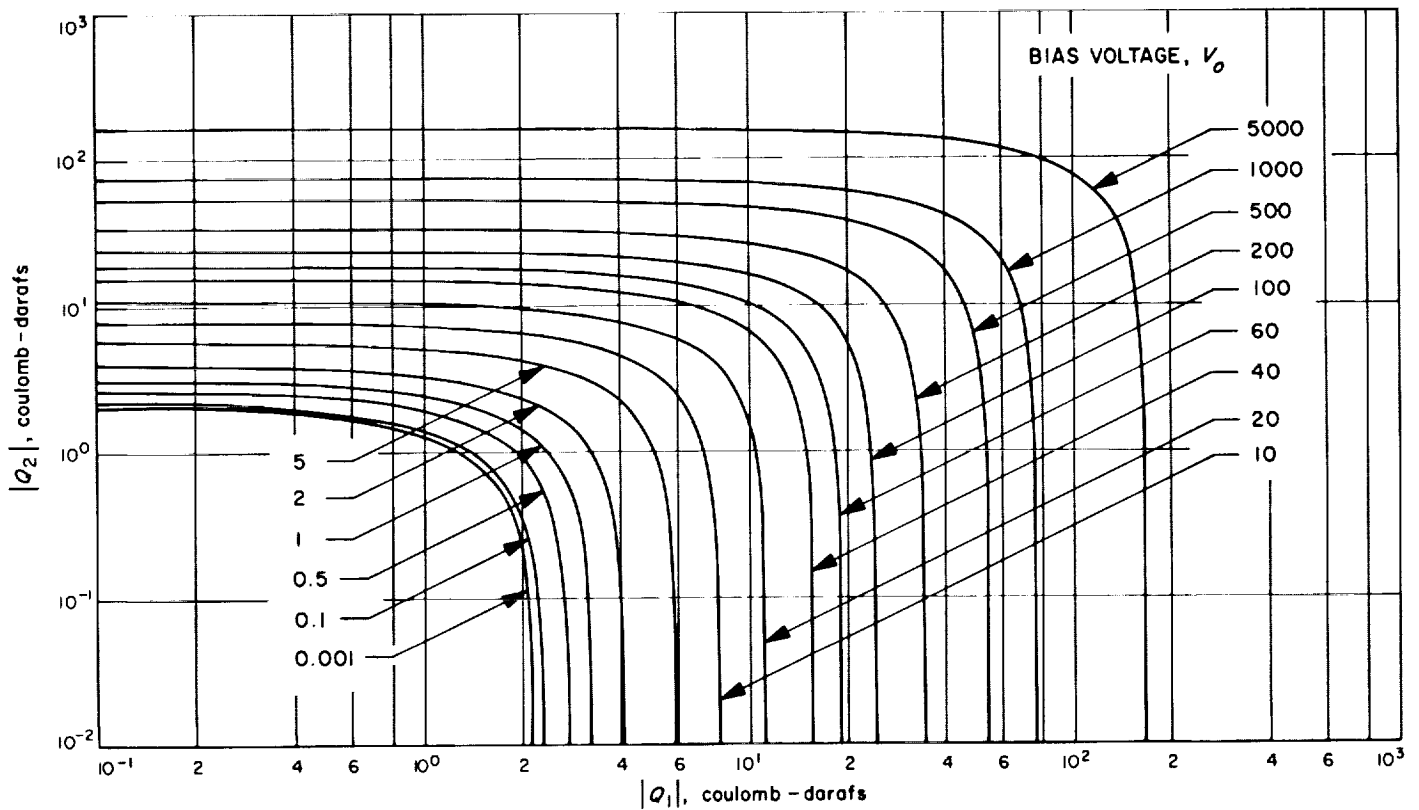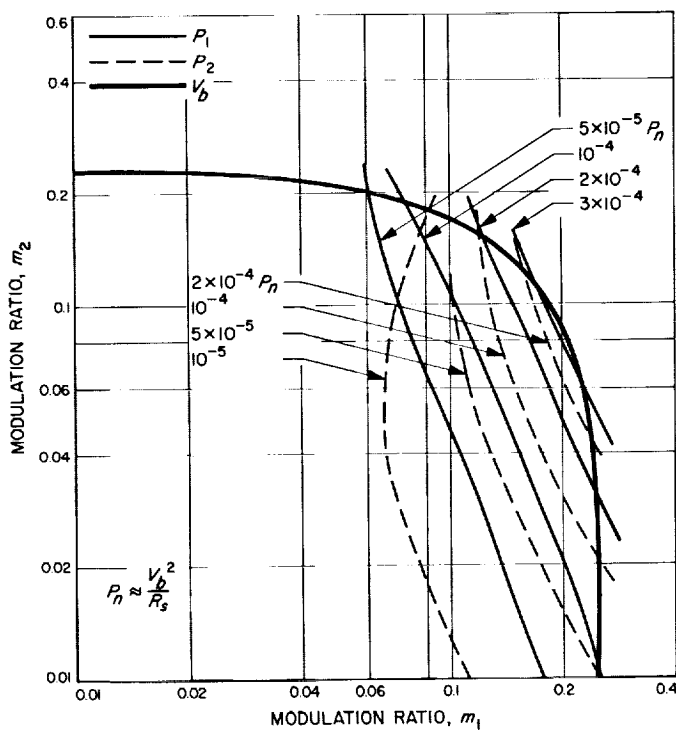
Fig. 32. Bias voltage

Fig. 33. Input and output powers of abrupt–junction doubler
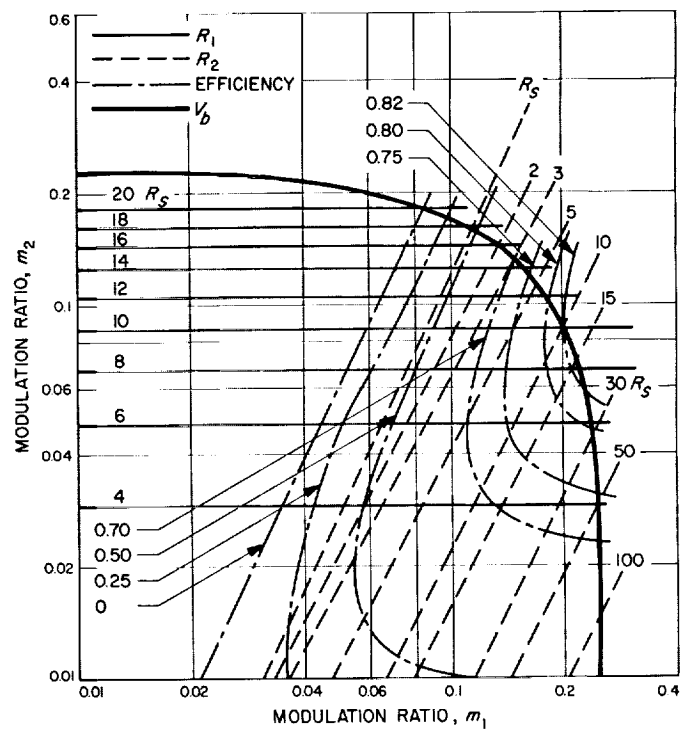

Fig. 34. Input and load resistances and efficiency of abrupt-junction doubler

used since design tolerances can rarely be made less than 10%.

The hope is that the final result of this work will be a series of design curves for several frequencies for multiplication and division by rational numbers useful in communication systems.

C. Radar Astronomy

W. F. Gillmore and D. D. Muhleman

1. Relationship Between Power Spectrum and Average Backscatter Function, *W. F. Gillmore*

a. Introduction. Power which is reflected from a surface, parallel to the incident power but reversed in direction, is called backscattered power. The intensity of this power may be found quantitatively from a backscatter function. This function is defined as 4π times the ratio of backscattered radiation intensity to the product of incident power density and surface area. In general, a backscatter function depends in a very complicated way on the orientation of the scattering surface.

It is a useful simplification to define an average backscatter function. This function F is assumed to depend only upon the polar angle θ measured from the surface normal to the direction of the incident signal. Each infinitesimal element of area on any given surface is defined to have the same average backscatter function. The total amount of backscattered power is equal to the sum of the power backscattered from each elemental area.

The average backscatter function for the Moon or a planet can be found from the power spectral density of its radar echo. A simple numerical method for doing this has been previously described by Gillmore (Ref. 26). This method is based upon two transform equations of the form

$$P(s) = \frac{1}{a} \int_{\sin^{-1}s}^{\pi/2} \frac{F(\theta) \sin \theta \, d\theta}{(\sin^2 \theta - s^2)^{1/2}} \tag{1}$$

$$F(\theta) = -\frac{2a \cos \theta}{\pi} \int_{\sin \theta}^{1} \frac{P'(s) \, ds}{(s^2 - \sin^2 \theta)^{1/2}} \tag{2}$$

which relate the power spectrum $P(s)$ to the average backscatter function $F(\theta)$. The variable s is a normalized frequency variable with a magnitude less than or equal to unity while a is the half bandwidth in cycles per second of $P(s)$. The average backscatter function F depends upon the angle θ which ranges from 0 to $\pi/2$.

It is difficult to achieve a clear understanding of the significance of Eqs. (1) and (2) without knowing the manner in which they are obtained. Eq. (1) has a physical significance which will be discussed in the next section. Eq. (2) is the mathematical result of inverting Eq. (1). If both the physical and mathematical properties of these equations are used, many interesting facts can be deduced. The following parts of this article illustrate some methods which are useful for this purpose. A few examples are also given to help clarify the results.

b. Physical and mathematical properties. The radar echo from a rotating spherical target has a power spectral density which is determined by several things. In general, the shape is determined by the average backscatter function; the half bandwidth is determined by the rotation rate of the target, the radar frequency, and the angle between the axis of rotation and the radar signal. Vertical height of the power spectrum is determined by the relative strength of the received signal. Despite the many complicating factors, $P(s)$ will always have the normalized form given in Eq. (1).

Fig. 35 shows a simplified picture of the physical situation leading to Eq. (1). The loci of two strips of area having given doppler shifts are shown in the diagram. Approaching strips of area return power for positive values of S, while receding strips correspond to negative values of s. The leading and trailing edges of the planet correspond to $s = \pm 1$.

Experimentally it was found useful to average both halves of the power spectrum $P(s)$ together. This gives an improved estimate of the average $P(s)$ from the experimental data and makes it unnecessary to consider values of s outside the interval from 0 to 1. Eqs. (1) and (2) are actually already given in the correct form to apply to this case.

The total power illuminating a small element of surface area varies as the cosine of the angle θ between the surface normal and the direction of the incident radar signal. On a perfectly spherical planet the point in the center corresponds to a θ of zero, while the perimeter
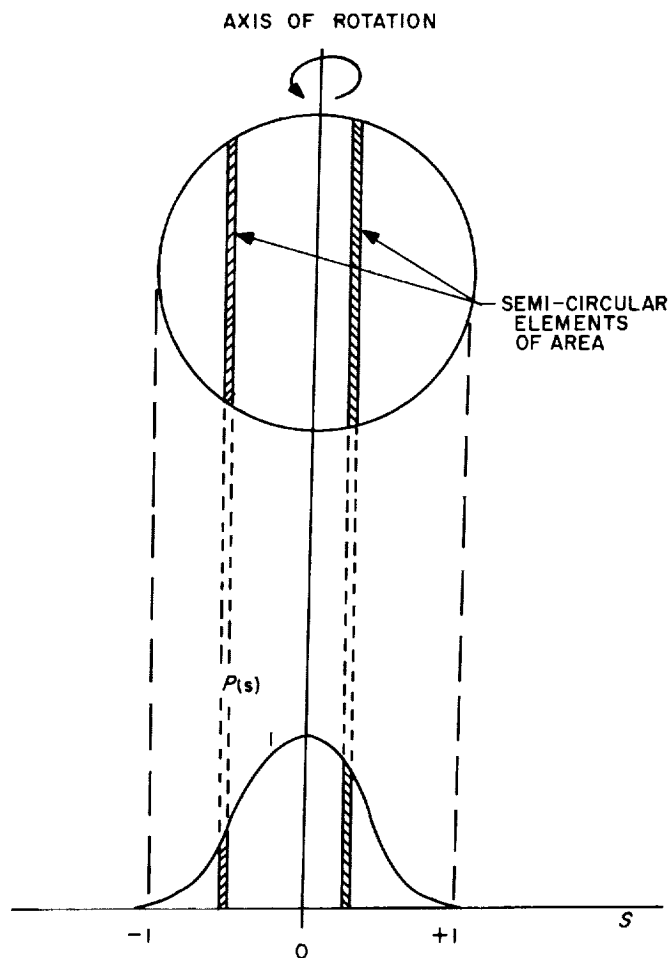
Fig. 35. Rotating planet and power spectrum of its radar echo

$$P''(s) = \frac{1 + s^2}{a(1 - s^2)^2} \int_{\sin^{-1}s}^{\pi/2} \frac{F'(\theta) \cos \theta \, d\theta}{(\sin^2 \theta - s^2)^{1/2}}$$

$$+ \frac{1}{a(1 - s^2)} \int_{\sin^{-1}s}^{\pi/2} \frac{F''(\theta) \sin \theta \, d\theta}{(\sin^2 \theta - s^2)^{1/2}} - \frac{F'(\pi/2)}{a(1 - s^2)^{3/2}}$$

$$(4)$$

$$F'(\theta) = \frac{2a \sin \theta}{\pi} \int_{\sin \theta}^{1} \frac{P'(s) - \text{ctn}^2(\theta) \, sP''(s)}{(s^2 - \sin^2 \theta)^{1/2}} \, ds$$

$$+ \frac{2a \cos \theta}{\pi \sin \theta} P'(1) \qquad (5)$$

$$F''(\theta) = \frac{2a \cos \theta}{\pi} \int_{\sin \theta}^{1} \frac{P'(s) \, ds}{(s^2 - \sin^2 \theta)^{1/2}}$$

$$+ \frac{6a \cos \theta}{\pi} \int_{\sin \theta}^{1} \frac{sP''(s) \, ds}{(s^2 - \sin^2 \theta)^{1/2}}$$

$$- \frac{2a \cos^3 \theta}{\pi \sin^2 \theta} \int_{\sin \theta}^{1} \frac{s^2 P'''(s) \, ds}{(s^2 - \sin^2 \theta)^{1/2}}$$

$$- \frac{2a(1 + \sin^2 \theta) P'(1)}{\pi \sin^2 \theta} + \frac{2a \cos^2 \theta \, P''(1)}{\pi \sin^2 \theta}$$

$$(6)$$

Using these equations together with the original transform pair [Eqs. (1) and (2)], it is possible to see exactly what happens at the end points of P and F. A sample of the results which can be obtained in this manner are shown in Table 3.

Some of the results in Table 3 follow immediately from the physical properties of $F(\theta)$, while others require additional mathematical restrictions. In most cases, sufficient conditions are given for the validity of each table entry. Two exceptions are $P''(1)$ and $F''(0)$ which are usually infinite. There are conditions under which these quantities can be finite, but it is somewhat unlikely that experimental data will satisfy them.

It is difficult to appreciate the significance of the results in Table 3 without looking at some specific examples. A few examples which were used to check some of these results are given in the next part of this article. The examples alone help to give a much better appreciation of the nature of the basic transform equations (1) and (2).

c. Some simple examples. All of the examples are listed in Table 4. The first column gives $P(s)$ and the second column gives $F(\theta)$. In order to make the equations easier to visualize, curves have been drawn for each entry in the table. Flat, linear, and parabolic shapes were chosen as about the simplest useful examples for

of the visible disk corresponds to the locus of points where θ is $\pi/2$. As θ approaches $\pi/2$, the incident power will go to zero. Since any real surface is never able to backscatter more power than that which falls upon it, the value of $F(\theta)$ is bounded above by the curve $\cos \theta$. It is also clear from the definition that F cannot be negative.

A physical property of having $F(\theta)$ bounded between zero and $\cos \theta$ leads to interesting conclusions. The power spectrum $P(s)$ is bounded and non-negative. It must also go to zero as s approaches unity.

Many other conclusions have been drawn by looking at the equations for the derivatives of $P(s)$ and $F(\theta)$. Equations for their first two derivatives are given below.

$$P'(s) = \frac{s}{a(1 - s^2)} \int_{\sin^{-1}s}^{\pi/2} \frac{F'(\theta) \cos \theta \, d\theta}{(\sin^2 \theta - s^2)^{1/2}} \qquad (3)$$

Table 3. End points and derivatives

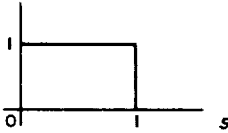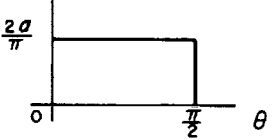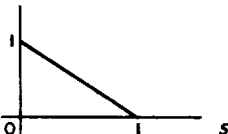| $P(s)$	$F(\theta)$
$P(0) = \dfrac{1}{a}\displaystyle\int_0^{\pi/2} F(\theta)\,d\theta$	$F(0) = -\dfrac{2a}{\pi}\displaystyle\int_0^1 \dfrac{P'(s)\,ds}{s}$
$P(1) = \begin{cases} \dfrac{\pi}{2a} F\!\left(\dfrac{\pi}{2}\right) \\ 0 \quad \text{if } |P'(1)| < \infty \end{cases}$	$F\!\left(\dfrac{\pi}{2}\right) = 0 \quad \text{if } |P'(1)| < \infty$
$P'(0) = 0 \quad \text{if } |F'(\theta)| < \infty$	$F'(0) = \begin{cases} -\infty & \text{if } P'(0) < 0 \\ 0 & \text{if } P'(0) = 0 \text{ and} \\ \left\| P''(1) - \displaystyle\int_0^1 \dfrac{P'''(s)\,ds}{(s^2 - \sin^2\theta)^{\frac{1}{2}}} \right\| < \infty \end{cases}$
$P'(1) = \begin{cases} -\infty & \text{if } F'\!\left(\dfrac{\pi}{2}\right) < 0 \\ \text{finite} & \text{if } F\!\left(\dfrac{\pi}{2}\right) = F'\!\left(\dfrac{\pi}{2}\right) = 0 \end{cases}$	$F'\!\left(\dfrac{\pi}{2}\right) = 0 \quad \text{if } |P'(1)| < \infty$
$P''(0) = \dfrac{1}{a}\displaystyle\int_0^{\pi/2} F'(\theta)\,\text{ctn}\,\theta\,d\theta - \dfrac{F'(0)}{a}$	$F''(0) = -\infty \quad \text{usually}$
$P''(1) = \infty \quad \text{usually}$	$F''\!\left(\dfrac{\pi}{2}\right) = -\dfrac{4a}{\pi} P'(1)$

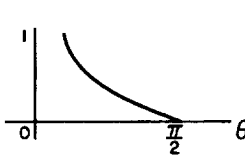**Table 4. Simple backscatter functions**
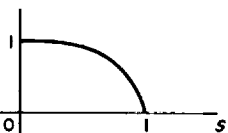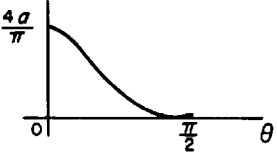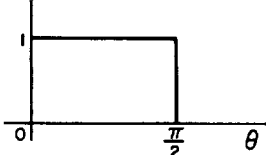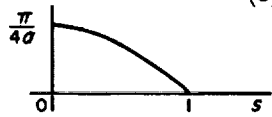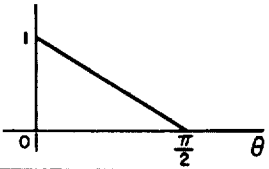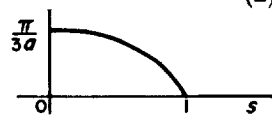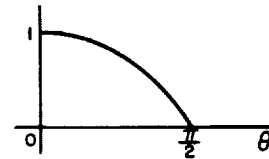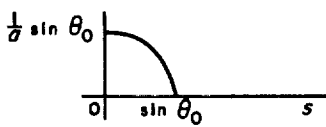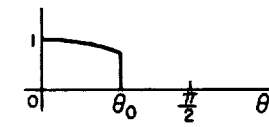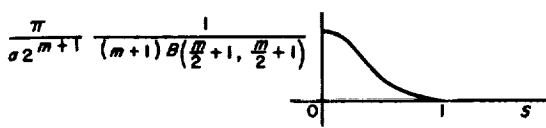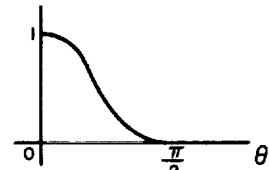
$P(s)$	$F(\theta)$
1	$\dfrac{2a}{\pi}$
$1 - s$	$\dfrac{a}{\pi}\cos\theta \log\!\left[\dfrac{1 + \cos\theta}{1 - \cos\theta}\right]$
$1 - s^2$	$\dfrac{4a\cos^2\theta}{\pi}$
$\dfrac{\pi}{2a}$	1

Table 4. Simple backscatter functions (cont'd)

$P(s)$	$F(\theta)$
$\dfrac{2}{\pi a}\displaystyle\int_0^{(1-s^2)^{1/2}}\dfrac{1}{x}\tanh^{-1}x\,dx = \dfrac{2}{\pi a}\left[(1-s^2)^{1/2} \right.$ $\left. +\dfrac{(1-s^2)^{3/2}}{(3)^2}+\cdots\right]$	$1-\dfrac{2}{\pi}\theta$
$\dfrac{4}{\pi a}\displaystyle\int_0^{(1-s^2)^{1/2}}\dfrac{1}{x}\ln(1+x)\,dx = \dfrac{4}{\pi a}\left[(1-s^2)^{1/2}\right.$ $\left. -\dfrac{(1-s^2)^{3/2}}{(2)^2}+\cdots\right]$	$1-\dfrac{4}{\pi^2}\theta^2$
$\begin{cases}\dfrac{1}{a}(\sin^2\theta_0-s^2)^{1/2} & \text{if } 0\le s<\sin\theta_0 \\[2mm] 0 & \text{if } \sin\theta_0\le s\le\dfrac{\pi}{2}\end{cases}$	$\begin{cases}\cos\theta & \text{if } 0\le\theta<\theta_0 \\[2mm] 0 & \text{if } \theta_0<\theta\le\dfrac{\pi}{2}\end{cases}$
$\dfrac{\pi}{a2^{m+1}}\dfrac{(1-s^2)^{m/2}}{(m+1)B\left(\dfrac{m}{2}+1,\dfrac{m}{2}+1\right)}$	$\cos^m\theta$

$P(s)$ and $F(\theta)$. The last two entries in the table are given because of their rather general form and the ease with which they can be obtained analytically.

Lines 1, 2, and 4 of the table illustrate power spectra which correspond to physically impossible backscatter functions. Each of the other table entries are properly bounded. These examples show that if $F(\theta)$ is monotonically decreasing, $P(s)$ is also monotonically decreasing. This result follows from Eq. (3). On lines 1, 4, 7 and 8, $F(\theta)$ is flat over part or all of its range. In each of these cases there is a corresponding flat region in the power spectrum.

The example shown on the seventh line of the table gives a power spectrum for a truncated cosine type backscatter function. This relation can be deduced directly from basic principles as described in the last part of this article. Using this together with the linear property of the backscatter process, it is possible to derive the transform pair [Eqs. (1) and (2)]. This derivation is quite elementary and makes use of certain superposition integrals.

At the end of the table is the cosine power type backscatter function. This relation has been used quite extensively and has an interesting history. It is perhaps

the easiest general relation that can be found between $P(s)$ and $F(\theta)$. Before better methods were developed, it was used as a means of finding backscatter functions.

In one approach the independent variable of the power spectrum was first changed by letting $s = \sin \phi$. The new curve $P(\phi)$ was expanded in a Fourier series containing terms of the form $\cos n\phi$. If the series converged rapidly enough, the first few terms could be converted into a sum of terms of the form $\cos^m \phi = (1 - \sin^2 \phi)^{m/2} = (1 - s^2)^{m/2}$. Coefficients of these terms were then used with the transformation given in the last line of the table. The result was a sum of terms of the form $\cos^m \theta$. This was used as an approximation to the desired backscatter function.

It is not difficult to see the many problems which are inherent in this double approximation method. Despite these drawbacks the cosine power backscatter functions have been used both as an indication of the shape of experimental data and as a form for empirical back-scattering functions. Obviously there is much to be gained by using the exact methods based on Eqs. (1) and (2).

One of the things which tremendously simplifies work with backscatter functions is the ability to go directly from $F(\theta)$ to $P(s)$. This allows one to estimate the accuracy of a complicated average backscatter function or check the validity of a new empirical backscattering law. There is a very convenient computer program which performs this operation and plots both input and output results. This program is described in d which follows.

d. A computer program for finding P(s). If the average backscatter function $F(\theta)$ is known, this program will give $P(s)$ and plot both $F(\theta)$ and $P(s)$. When $F(\theta)$ is found by another program described by Gillmore (Ref. 26), the results are in a form suitable for input to this program. It should be noted, however, that the reverse is not true. This is because $F(\theta)$ is normally computed from an experimentally measured autocorrelation function rather than from the power spectrum $P(s)$.

The program uses an integration which is equivalent to that shown in Eq. (1). This equation is not suitable for numerical integration as it stands because the denominator of the integrand vanishes at the lower limit. By substituting U for the square root in the denominator, the equation takes on the form shown in Eq. (7), which can be solved numerically and has been found to give very good results.

$$P(s) = \frac{1}{a} \int_0^{(1-s^2)^{1/2}} \frac{F[\sin^{-1}(U^2 + s^2)^{1/2}]\,du}{(1 - s^2 - U^2)^{1/2}} \qquad (7)$$

The denominator of the integrand in Eq. (7) goes to zero at the upper limit. Fortunately, the numerator is also zero because of the way $F(\theta)$ is bounded. When the integrand is evaluated, it is found to be $-F'(\pi/2)$. Since any experimentally measured power spectrum must necessarily have $|P'(1)| < \infty$, it follows from Table 3 that the integrand at the upper limit is zero.

In rare cases it may be useful to evaluate $P(s)$ when $F'(\pi/2) \neq 0$. This can be accomplished by inserting the proper value of $-F'(\pi/2)$ into the program. Because of the way $F(\theta)$ is bounded, it follows that $-F'(\pi/2)$ must be between zero and one. Thus, the improvement obtained by making the integrand exact at the upper limit may be so small that it is not worthwhile.

Since the integrand in Eq. (7) is relatively smooth, Simpson's rule works quite well as the numerical integration formula. It requires that the integrand be evaluated at a number of equally spaced points. Equal increments in U correspond to unequal increments in the argument of F. Values of F at these points are computed by linear interpolation between the known points which are usually spaced at increments of 1 deg. Linear interpolation has given excellent results on average backscatter functions computed from even the noisiest experimental data. There has been no need to go to more intricate interpolation schemes. Calculation of the denominator of the integrand in Eq. (7) introduces no special difficulties.

Fig. 36 shows a sample of the plotted output from this program. The input average backscatter function is normalized so that its maximum value is 1.0. Its abscissa runs from 0 to 90 deg. Since this curve is mostly near the left side of the plot, the computed power spectrum $P(s)$ was centered at 50 on the abscissa. Values of 100 and 0 correspond to $s = \pm 1$. Both halves of the $P(s)$ curve were plotted to make it look more symmetrical, but both halves of the curve are identical. The peak value of $P(s)$ is normalized to unity for convenience.

This program is very fast. The running time on any computer is roughly proportional to the product of the number of points in the computed power spectrum $P(s)$ and the number of steps used in the numerical integration. Neglecting the small but fixed input–output time, the proportionality constant for the 7094 was less than 3 msec per point per step.

e. Conclusions. Although the theory and computational schemes discussed here were originally developed for

the purpose of checking computed average backscatter functions, they are finding much wider application. They serve as a guide for anyone wishing to work on empirical or semi-empirical scattering theories. Also they can serve as a valuable tool for those who are interested in devel-

oping more exotic mathematical methods for computing average backscatter functions. It is useful to have an accurate idea of the average backscatter function of any lunar or planetary surface in order to be able to tell which irregularities in $P(s)$ are due to actual surface

Fig. 36. Sample computer output with $F(\theta) = \cos^{25}\theta$ input
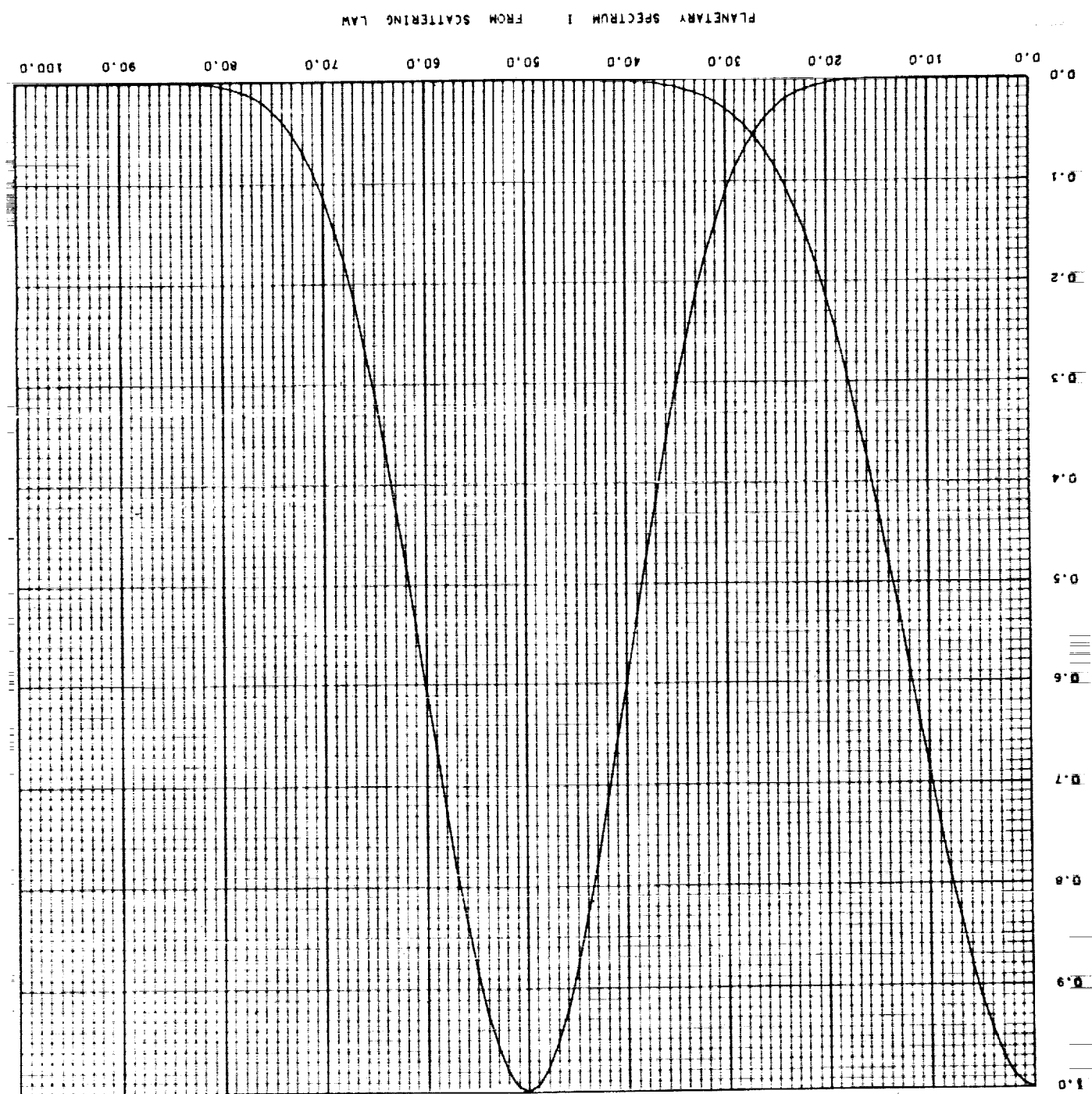
features. Thus, this work represents a small step forward toward the ultimate goal of providing a method for mapping the Moon and the planets by radar.

2. On the Venus Microwave Phase Effect,
D. O. Muhleman

a. *Introduction.* According to the measurements of Kuzmin and Solomonovich (Refs. 27, 28) the 8-mm brightness temperature of Venus exhibits a relatively strong variation with the planetary phase angle. This effect has been observed further by Mayer (Ref. 29) at 3 cm and by Drake (Ref. 30) at 10 cm, among others. Generally, the observed effect is that the brightness temperature reaches a minimum value near inferior conjunction (phase of 180 deg). Considerable doubt concerning the physical reality of this phenomenon has been expressed by many investigators, apparently because of the well-known absence of a phase effect in the infrared domain (Murray, Ref. 31). Indeed, it is easy to hypothesize the existence of systematic errors in the radio observations which would cause such an inference in the data. In this report, a reasonable theory is developed for the origin of such a physical phenomenon and thereby credence is added to its reality.

The microwave observations of Venus in terms of the amplitude of the temperature variations and the phase shifts of the variations are similar to the microwave observations of the Moon. The lunar temperature curves are well understood in terms of the theory developed by Piddington and Minnett (Ref. 32) and the applications of this theory by Troitsky (Ref. 33).

b. *Theory of thermal microwave emission.* If the *surface* temperature of a semi-infinite homogeneous slab, T_s, is a periodic function of time t, with a fundamental period $2\pi/\Omega$, i.e.,

$$T_s = T_0 + \sum_{n=1}^{\infty} T_n \cos n\Omega t \tag{1}$$

then (according to Carslow and Jaeger, Ref. 34) the temperature at a depth X is given by

$$T(X) = T_0 + \sum_{n=1}^{\infty} T_n e^{-\beta_n X} \cos(n\Omega t - \beta_n X) \tag{2}$$

where β_n is the thermal attenuation coefficient

$$\beta_n = \left(\frac{n\Omega\rho c}{2k}\right)^{\frac{1}{2}} \tag{3}$$

and k is the thermal conductivity, ρ the density, and c the specific heat of the material. Eq. (2) represents a series of temperature waves traveling into the material, the harmonics diminishing in amplitude and shifting in phase more rapidly than the fundamental. The brightness temperature observed at a given microwave length will consequently depend on the opacity of the material (electrical parameters) which effectively determines the depth of emission at a given wavelength. Piddington and Minnett (Ref. 33) have solved the equation of transfer for this case and obtain the brightness temperature

$$T_b = T_0 + \frac{T_1}{[1 + 2\delta + 2\delta^2]^{\frac{1}{2}}} \cos(\Omega t - \psi) \tag{4}$$

where

$$\delta = \frac{\beta_1}{K} \text{ (for normal incidence)} \tag{5}$$

$$\tan\psi = \frac{\delta}{1 + \delta} \tag{6}$$

and K is the (radio) volume absorption coefficient, normally a function of wavelength, λ. Thus, we see from Eq. (4) that the observed brightness temperature at a given wavelength would have a mean of T_0 with a variable amplitude and phase shift determined by the electrical and thermal properties of the body. Clearly, the phase shift arises because the temperature of a deep layer cools (or warms) to a minimum (or maximum) value later in time than the layers above it. In general, the longer wavelength radiation arises from the deeper layers.

In dielectrics of relatively low electrical conductivity the absorption coefficient K is given by

$$K \simeq \frac{4\pi\sigma}{C_0 \varepsilon} \quad \text{(cgs units)} \tag{7}$$

where σ is the electrical conductivity, C_0 the vacuum speed of light, and ε the specific dielectric constant of the material. K can be expressed in the form of the loss tangent which is more convenient for our purposes. Since by definition

$$\tan\Delta = \frac{2\sigma\lambda}{\varepsilon C_0}$$

then

$$K \simeq \frac{2\pi(\varepsilon)^{\frac{1}{2}}}{\lambda} \tan\Delta. \tag{8}$$

It is clear from Eq. (4) that the quantity of interest which can be obtained from the observations is δ. The theoretical value of δ from Eqs. (3), (5), and (8) is .

$$\delta = \frac{\lambda \left(\frac{\Omega \rho c}{2k}\right)^{1/2}}{2\pi \, (\epsilon)^{1/2} \tan \Delta} \qquad (9)$$

Thus, from Eq. (9), the quantity δ/λ will be a constant for a homogeneous body within a wavelength range for which $\tan \Delta$ is wavelength independent. Troitsky (Ref. 33) in particular, has demonstrated that for the Moon $\delta/\lambda \simeq 2$ for a wavelength range from 8 mm to 10 cm which from Eq. (9) implies the constancy of $\tan \Delta$. The result is also established from the wealth of lunar radar observations. It should be pointed out that the loss tangents for most Earth soils remain constant over a considerable range of microwave frequencies.

c. Discussion of the Venus observations. Brightness temperature measurements of Venus with sufficient sensitivity and quantity to indicate a phase effect have been made at 10 cm (Drake, Ref. 30), 3.15 cm (Mayer, Ref. 29), and 8 mm (Kuzmin and Solomonovich, Ref. 28). The observations can best be expressed as the Fourier series fit in the form of a constant term and a fundamental of period $2\pi/\Omega_s$, where Ω_s is the synodic period of Venus:

10 cm: $T_b = 622\,(6) + 41\,(12) \cos\,[i - 21°\,(9)]$, °K

3.15 cm: $T_b = 621\,(7) + 73\,(9) \cos\,[i - 11.°7\,(33)]$, °K

8 mm: $T_b = 485\,(70) + 115 \cos\,[i + 2°\,(5)]$, °K

where $i = \Omega_s t$, and time is reckoned such that $i = 180$ deg at inferior conjunction. The uncertainties in the parentheses are taken from the papers referenced above except for the phase delay of the 8-mm data which the author has taken from the Soviet data.

The similarity in the 10- and 3.15-cm constant terms strongly suggests that the atmosphere of Venus has little effect on surface emission at these wavelengths, but it is well established that the "surface" emission at 8 mm is somewhat absorbed by the Venusian atmosphere. Based on the results of Spinrad (Ref. 34) concerning the surface pressure, we have taken the 8-mm normal opacity of $\tau = 0.1$ from Barrett's (Ref. 35) CO_2 and N_2 model atmosphere at 10 atmospheres pressure. This is, of course, uncertain but probably adequate for our purposes. Correcting the 8-mm variable component in Eq. (4), we obtain

$T_{v,0.8} = 127°K$. Then from Eq. (4) the ratio of the variable components at two different wavelengths is

$$\frac{T_{v1}}{T_{v2}} = \left(\frac{1 + 2\delta_2 + 2\delta_2^2}{1 + 2\delta_1 + 2\delta_1^2}\right)^{1/2} \qquad (10)$$

We now adopt the hypothesis for testing that

$$\frac{\delta_\lambda}{\lambda} = \alpha \qquad (11)$$

where α is constant. Using the 8-mm and 10-cm data where $T_{v,0.8}/T_{v,10} = 3.1$, the solution of Eq. (10) for α yields $\alpha = 0.2$. Carrying our hypothesis a step further, we can use the result $\delta = 0.2\lambda$ to predict the amplitude variation at 3.15 cm and all of the phase angles. The results are shown in Table 5. If the ratio of the observed values of the 3.15- and 10-cm temperature variations is used in Eq. (10), the resulting value of α is 0.193 and the results of Table 5 are essentially identical. The fifth column of Table 5 shows the theoretical phase shift (based on $\delta = 0.2\lambda$) advanced by 9.°8 to force the agreement with the 8-mm observations for which the phase at minimum temperature is most accurately known. If the 3.15- or 10-cm results are used to obtain a shift in phase, values of 9.°7 or 12.°7 are required, respectively. These values are all in agreement with the accuracy of the observations, and all suggest that the minimum surface temperature occurs before inferior conjunction. It is not surprising that such a shift is required because the theory of Piddington and Minnett was developed for a planet free of an atmosphere where the surface temperature is exactly in agreement with the phase relative to the Sun. In the case of Venus, one would expect the surface heating to be shifted by the presence of meteorological phenomena in the atmosphere just as on the Earth. The shift of 9.°8 apparently corresponds (using the Venusian synodic period) to 6 Earth days. However, the occurrence of minimum surface temperature before inferior conjunction is difficult to understand. One would expect the heating minimum to occur after inferior conjunction (analogously to the Earth)

Table 5. Comparison of theoretical and observed brightness temperature variations for Venus

λ	$T_v\,(\alpha = 0.2),$ °K	T_v (obs), °K	$\psi\,(\alpha = 0.2),$ deg	$(\psi - 9.°8)$	ψ (obs), deg
8 mm	(127)	127 ± ?	7.°8	(−2.°0)	−2° ±5
3.15 cm	(84)	73 ± 9	21.°4	11.°6	11.°7 ±33
10 cm	(41)	41 ± 12	33.°7	23.°9	21° ±9

due to the heat capacity of the atmosphere. This phenomenon could be caused by a flow of cold atmospheric gases from the cold side of the plant or a flow from the hot side. Kuzmin and Solomonovich (Ref. 28) have used the occurrence of observed 8-mm temperature minimum to infer a direct rotation of Venus but this appears highly speculative. It can be seen from Table 5 that the agreement between the theory (with the hypothesis, $\delta = 0.2\lambda$) and measurement is excellent. Clearly, an alternate approach to these calculations would be to compute a set of α's from all of the amplitude variations and phase delays, but this procedure leads to essentially the same result.

Thus, a simple model of a homogeneous body with thermal microwave emission from the surface appears to explain the Venusian observations to the accuracy of the measurements. If these results are applied at 21 cm, one gets an expected amplitude variation of 22°K and a phase shift of 39 deg. The phase shift suggests that the minimum 21-cm brightness temperature should occur about 20 days after inferior conjunction. This fact, coupled with the small amplitude variation, may explain why attempts to observe a 21-cm phase effect have been frustrating.

d. Interpretation of $\delta = 0.2\lambda$. If we accept $\delta = 0.2\lambda$ we get from Eq. (9) that

$$\frac{\left(\frac{\rho c}{2k}\right)^{\frac{1}{2}}}{2\pi\,(\epsilon)^{\frac{1}{2}}\tan\Delta} = 0.2. \qquad (12)$$

Reasonable values of $\rho \simeq 2$, $c \simeq 0.2$ can be obtained from terrestrial analogies, and according to Muhleman (Ref. 36), a value of $\epsilon \simeq 3.5$ has been obtained from Venus radar measurements. Troitsky (Ref. 33) reports that laboratory measurements of many pertinent terrestrial materials yield a consistently constant value of

$$\frac{\tan\Delta}{\rho} \simeq 10^{-2}.$$

A relationship between the thermal conductivity k and the synodic angular velocity Ω can be obtained using these values in Eq. (12). The result is

$$k = 138\Omega\left(\frac{\text{cal}}{\text{cm-sec °K}}\right) \qquad (13)$$

which is shown in graph form in Fig. 37 where k is plotted against $2\pi/\Omega$ expressed in Earth days. The nature of the curve suggests that this method is capable of yield-
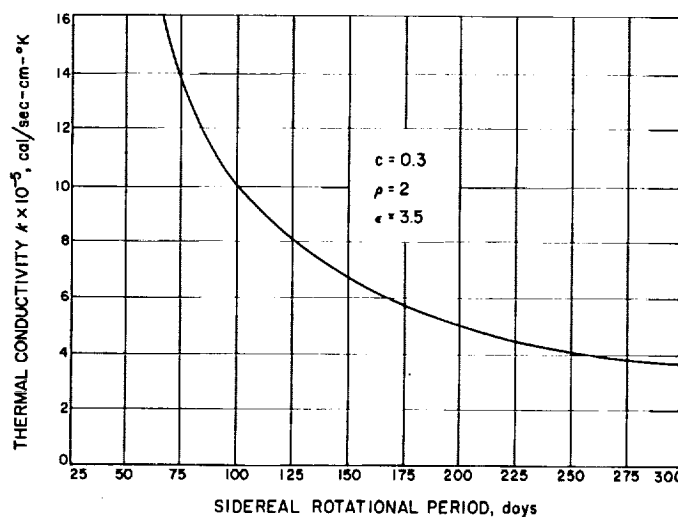

Fig. 37. Thermal conductivity versus Venusian rotational period consistent with $\delta = 0.2\lambda$

ing a good value of k if the rotational rate is known. A representative list of thermal conductivities for some terrestrial materials is shown in Table 6. These data indicate that a surface of dry terrestrial type sand with a rotational period in general agreement with the Venus radar results of Muhleman (Ref. 37), Goldstein (Ref. 38), and Carpenter (Ref. 39) yields a fair model of the Venusian surface. These results argue against the existence of basalt, granite, and other igneous rocks (loosely speaking) as major constituents of the Venusian surface and in favor of granulated forms such as sand. Of course, a wide range of materials would exhibit values of thermal conductivity and dielectric constant consistent with those reported here if the materials occur in a granulated state with packing factors considerably less than unity.

The values of the thermal conductivity shown in Fig. 37 are clearly low relative to most natural terrestrial

Table 6. Thermal conductivity and dielectric constant for some terrestrial materials

Material	Thermal conductivity, k, $\dfrac{\text{cal}}{\text{cm sec °K}}$	Dielectric constant, ϵ, at $T \sim 290°K$
Dry sand	6×10^{-4}	2–4
Basalt	4.8×10^{-3}	7–17
Granite	5.3×10^{-3}	7
Sandstone	3×10^{-4}	4.8
Graphite	$4 \rightarrow 9 \times 10^{-4}$	—
Mica	1.2×10^{-3}	5.5–10
Petroleum	3×10^{-4}	1.5–3.0

materials. These values are strongly dependent on our choice of ρ, c, and ε used in the computations. Little information is available concerning even the terrestrial materials at temperature as high as 600°K. In general, one would expect the heat capacity, c, to increase with temperature in this region yielding significantly higher values of k. Furthermore, the thermal conductivities of the terrestrial samples tend to decrease linearly with an increase in temperature. The thermal inertia defined by $(k\rho c)^{-\frac{1}{2}}$ is somewhat better separated in our calculations which yield values of 205 and 290 for a Venusian rotational period of 100 and 200 days, respectively. These values are consistent with granular or cellular terrestrial substances. Further analysis of this problem should include a detailed study of pertinent chemical equilibria reactions, see for example, Mueller (Ref. 40). Perhaps it should be noted in passing that the existence of hydrocarbons on the Venusian surface is not ruled out by these results. Some confusion remains concerning the direction of the Venusian axial rotation. The apparent occurrence of the surface temperature minimum before inferior conjunction suggests a direct rotation. However, the *sequence* of occurrence of the radio brightness minima at increas-

ing phase angles after inferior conjunction with increasing wavelengths more strongly suggests retrograde rotation.

e. Summary. It appears that the observed microwave brightness temperature variations with planetary phase can be explained in terms of the blackbody emission at varying depths from a homogeneous body. This would indicate that the atmosphere of Venus has relatively little effect on the thermal radio emission in the wavelength range from 8 mm to 10 cm. This result is consistent with the results of Muhleman (Ref. 36) concerning the transparency of the Venusian atmosphere at radar wavelengths of 12.5 and 68 cm.

The analysis of this paper indicates that the loss tangent of the Venusian surface material is relatively constant at the wavelengths studied (which is again consistent with the constancy of the surface reflectivity for decimeter and decameter radar waves). The analysis supports the idea of a very slow Venusian rotational rate with period of from 100 to 250 days and leads to a value of the thermal conductivity of the Venusian surface material similar to dry terrestrial materials.

References

1. Posner, E. C., "Effectiveness of Error-Correcting Codes," *Research Summary No. 36-13*, pp. 32–38, Jet Propulsion Laboratory, Pasadena, California, March 1, 1962.

2. Posner, E. C., "Coding Theory," *Space Programs Summary No. 37-19*, Vol. IV, pp. 150–151, Jet Propulsion Laboratory, Pasadena, California, February 28, 1963.

3. Weintraub, S., "Tables of the Cumulative Binomial Probability Distribution for Small Values of p," Free Press of Glencoe, New York, 1963.

4. Peterson, W. W., "Error-Correcting Codes," Wiley & Sons, New York, 1961.

5. Solomon, G., "Linear Recursive Sequences as Finite-Difference Equations," Group Report 47-37, Lincoln Laboratory, Lexington, Massachusetts, 1960.

6. Block, R. E., "Difference Sets, Block Designs, and Constant Distance Codes," *Space Programs Summary No. 37-22*, Vol. IV, pp. 137–138, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.

7. Titsworth, R. C., "Binary Cyclic Constant-Distance Codes," *Space Programs Summary No. 37-22*, Vol. IV, pp. 147–153, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.

8. Block, R. E., "Groups of Collineations on Codes," *Space Programs Summary No. 37-25*, Vol. IV, pp. 158–160, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

References (Cont'd)

9. Block, R. E., "Hadamard Matrices and Constant-Distance Codes," *Space Programs Summary No. 37-23*, Vol. IV, pp. 156–158, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

10. Viterbi, A. J., "Phase-Locked Loop Dynamics in the Presence of Noise by Fokker-Planck Techniques," Technical Report No. 32-427, Jet Propulsion Laboratory, Pasadena, California, March 1963.

11. Viterbi, A. J., "Optimum Coherent Demodulation for Continuous Modulation Systems," *Proceedings of the National Electronics Conference*, Vol. 18, pp. 498–508, Chicago, Illinois, October 1962.

12. Lehan, F. W., and Parks, R. J., "Optimum Demodulation," *IRE Convention Record*, part 8, pp. 101–103, PGIT, March 1953.

13. Davenport, W. B., Jr., "Signal-to-Noise Ratios in Bandpass Limiters," *Journal of Applied Physics*, Vol. 24, No. 6, pp. 720–727, June 1953.

14. Middleton, D., *Introduction to Statistical Communication Theory*, McGraw-Hill Book Co., New York, 1960.

15. Jaffe, R. M., and Rechtin, E., "Design and Performance of Phase Lock Circuits Capable of Near-Optimum Performance over a Wide Range of Input Signals and Noise Levels," *IRE Transactions on Information Theory*, Vol. IT-3, pp. 135–144, March 1957.

16. Stiffler, J. J., "On the Allocation of Power in a Synchronous Binary Communication System," *Space Programs Summary No. 37-25*, Vol. IV, pp. 176–181, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

17. Van Trees, H. L., "Synchronization and Bit Timing," *IEEE Transactions on Communications Systems*, December 1963.

18. Davenport, W. B., Jr., and Root, W. L., *Random Signals and Noise*, McGraw-Hill, New York, 1958.

19. Kac, M., and Siegert, A. J. F., "On the Theory of Noise in Radio Receivers with Square-Law Detectors," *Journal of Applied Physics*, Vol. 18, April 1947.

20. Ronen, Y., and Zakai, M., "The Maximum Likelihood Estimator for a Phase Comparison Angle Measuring System," *Proceedings of the IEEE*, No. 51, p. 1669, November 1963.

21. Rice, S. O., "Mathematical Analysis of Random Noise." Appears in Wax, N., *Selected Papers on Noise and Stochastic Processes*, p. 133, Dover Publications, New York, 1954. See in particular pp. 238–240.

22. Price, R., "Error Probabilities for Adaptive Multichannel Reception of Binary Signals," *IRE Transactions on Information Theory*, IT-8, p. S305, Eq. (8), September 1962.

23. Marcum, J. I., *A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix*, The RAND Corporation, Research Memorandum RM-773, July 1, 1948, p. 5. Also appears in *IRE Transactions on Information Theory*, IT-6, p. 159, April 1960.

References (Cont'd)

24. Penfield, P., and Rafuse, R. P., *Varactor Applications*, The MIT Press, Cambridge, Massachusetts, 1962.

25. Luettgenau, G., Williams, J., and Miyahira, H., "A Practical Approach to the Design of Parametric Frequency Multipliers, Bulletin No. 92-562, Pacific Semiconductors, Inc., California.

26. Gillmore, W. F., "A Numerical Method for Finding Backscatter Functions," *Space Programs Summary No. 37-23*, Vol. IV., pp. 188–191, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

27. Kuzmin, A. D., and Solomonovich, A. E., "Radio Observations of Venus in 1961," *USSR Astronomicheskii Zhurnal*, Vol. 5, p. 851, 1961.

28. Kuzmin, A. D., and Solomonovich, A. E., "Some Conclusions About Physical Conditions on Venus According to Radio Astronomical Observations at P. N. Lebedev Physical Institute," *Physics of the Planets*; Proceedings of the 11th International Astrophysical Symposium, Liège, Belgium, 1962.

29. Mayer, C. H., McCullough, T. P., Sloanaker, R. M., "3.15-cm Observations of Venus in 1961," *Physics of the Planets*, Proceedings of the 11th International Astrophysical Symposium, Liège, Belgium, 1962.

30. Drake, F. D., "Microwave Observations of Venus, 1962–1963," *Astronomical Journal*, Vol. 69, p. 62, 1962.

31. Murray, B. C., Wileg, R. L., Westphal, J. A., 'Infrared Photometric Mapping of Venus through the 8- to 14-Micron Atmospheric Window," *Journal of Geophysical Research*, Vol. 68, p. 4813, 1963.

32. Piddington, J. H., and Minnett, H. C., "Microwave Thermal Radiation from the Moon," *Australian Journal of Science*, Vol. A2, p. 63, 1949.

33. Troitsky, V. S., "Radio Emission of the Moon, Its Physical State, and the Nature of Its Surface," *The Moon*, edited by Kopal and Mikhailov, Academic Press, 1962.

34. Spinrad, H., "Spectroscopic Temperature and Pressure Measurements in the Venus Atmosphere," *Publications of the Astronomical Society of the Pacific*, Vol. 74, p. 187, 1962.

35. Barrett, A. H., "Microwave Absorption and Emission in the Atmosphere of Venus," *Astrophysical Journal*, Vol. 133, p. 281, 1961.

36. Muhleman, D. O., "The Electrical Characteristics of the Atmosphere and Surface of Venus from Radar Observations," *Icarus*, Vol. 1, p. 401, 1963.

37. Muhleman, D. O., "Radar Scattering from Venus and the Moon," *Astronomical Journal*, Vol. 69, p. 34, 1964.

38. Goldstein, R., "Venus Characteristics by Earth Based Radar," *Astronomical Journal*, Vol. 69, p. 12, 1964.

39. Carpenter, R., "Study of Venus by CW Radar," *Astronomical Journal*, Vol. 69, p. 2, 1964.

40. Mueller, R. F., "Chemistry and Petrology of Venus: Preliminary Deductions," *Science*, Vol. 141, p. 1046, 1963.